



GERMAIN & MAUREAU

Fondé en 1849
Conseils en Propriété Industrielle
European Patent Trademark & Design Attorneys

12, RUE BOILEAU
F-69006 LYON
BP 6153
F-69466 LYON CEDEX 06
TEL: (33) 04 72 69 84 30
FAX: (33) 04 72 69 84 31

The below-named translator is proficient in the French and English languages and verifies that the attached is a substantially correct English translation of the enclosed French language document.



Edith GAGNEUR
Translator to GERMAIN & MAUREAU

10 September 2009

Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS) of which the complete cause still remains unknown.

5

Numerous studies have supported the hypothesis for a viral etiology of the disease, but none of the known viruses tested has proved to be the causative agent tested for: a review of the viruses tested for in MS for many years has been carried out by E. Norrby and R.T. Johnson.

Recently, a retrovirus, different from the known human retroviruses, was isolated from patients suffering from MS. The authors were able to show that this retrovirus could be transmitted in vitro, that patients suffering from MS produced antibodies capable of recognizing proteins associated with the infection of the leptomeningeal cells by this retrovirus, and that the expression of the latter could be greatly stimulated by the immediate-early genes of some herpesviruses.

All these results argue in favor of the role in MS of at least one unknown retrovirus or of a virus having a reverse transcriptase (RT) activity which is detectable by the method published by H. Perron and termed "LM7-type RT" activity.

The studies by the applicant have made it possible to obtain two continuous cell lines infected with natural isolates obtained from two different patients suffering from MS, by a culture method as described in the document WO-A-93 20188, whose content is incorporated by reference into the present description. These two lines derived from cells of human choroid plexus, called LM7PC and PLI-2, were deposited at the E.C.A.C.C. on 22 July 1992 and 8 January 1993, respectively, under numbers 92 072201 and 93 010817, in accordance with the provisions of the Treaty of Budapest. Moreover, the viral isolates possessing an

LM7-type RT activity have also been deposited at the E.C.A.C.C. under the overall name of "strains". The "strain" or isolate harbored by the PLI-2 line, called POL-2, was deposited at the E.C.A.C.C. on 22 July 1992 under No. V92072202. The "strain" or isolate harbored by the LM7PC line, called MS7PG, was deposited at the E.C.A.C.C. on 8 January 1993 under No. V93010816.

Using the above-mentioned cultures and isolates, characterized by biological and morphological criteria, efforts were then made to characterize the genetic material associated with the viral particles produced in these cultures.

The proportions of genome already characterized were used to develop molecular detection tests for the viral genome and immunoserological tests, using the amino acid sequences encoded by the nucleotide sequences of the viral genome, in order to detect the immune response directed against epitopes associated with the viral infection and/or expression.

These tools have already made it possible to confirm an association between MS and the expression of the sequences identified in the patents cited further on. However, the viral system discovered by the applicant is related to a complex retroviral system. Indeed, the sequences which are found to be encapsidated in the extracellular viral particles produced by the different cultures of cells of patients suffering from MS show clearly that there is co encapsidation of retroviral genomes which are related but different from the "wild-type" retroviral genome which produces the infectious viral particles. This phenomenon was observed between replicative retroviruses and endogenous retroviruses belonging to the same family, or even heterologous retroviruses. The concept of endogenous retrovirus is very important in the context of our discovery because, in the case of MSRV-1, it has been

observed that endogenous retroviral sequences comprising sequences homologous to the MSRV-1 genome exist in normal human DNA. The existence of endogenous retroviral elements (ERV) related to MSRV-1 through all or part of their genome explains the fact that the expression of the MSRV-1 retrovirus in human cells can interact with related endogenous sequences. These interactions are found in the case of pathogenic and/or infectious endogenous retroviruses (for example some ecotropic strains of the Murine Leukemia virus), in the case of exogenous retroviruses whose nucleotide sequence may be found partially or completely in the form of ERVs, in the genome of the host animal (e.g. mouse mammary tumor exogenous virus transmitted via milk). These interactions consist mainly of (i) a transactivation or co-activation of ERVs by the replicative retrovirus, (ii) an "illegitimate" encapsidation of related RNAs of ERVs, or of ERVs - or even of cellular RNAs - simply possessing compatible encapsidation sequences, into the retroviral particles produced by the expression of the replicative strain, which are sometimes transmissible and sometimes with an inherent pathogenicity, and (iii) relatively high recombinations between the co-encapsidated genomes, in particular in the reverse transcription phases, which lead to the formation of hybrid genomes, which are sometimes transmissible and sometimes with an inherent pathogenicity.

Thus, (i) various MSRV-1-related sequences have been found in purified viral particles; (ii) molecular analysis of the various regions of the MSRV-1 retroviral genome should be carried out by systematically analyzing the co-encapsidated, interfering and/or recombinant sequences which are generated by the infection and/or expression of MSRV-1; furthermore, some clones may have portions of defective sequences produced by the retroviral replication and the template and/or transcription errors

caused by reverse transcriptase; (iii) the families of sequences related to the same retroviral genomic region are the supports for an overall diagnostic detection which may be optimized by the identification of invariable regions among the clones expressed and by the identification of reading frames responsible for the production of antigenic and/or pathogenic polypeptides which may only be produced by a portion, or even only one, of the clones expressed and under these conditions, the systematic analysis of the clones expressed in one region of a given gene makes it possible to evaluate the frequency of variation and/or recombination of the MSRV-1 genome in this region and to define the optimum sequences for the applications, in particular the diagnostic applications; (iv) the pathology caused by a retrovirus such as MRSV-1 may be a direct effect of its expression and of the proteins or peptides produced as a result, but also an effect of the activation, encapsidation, recombination of related or heterologous genomes and proteins or peptides produced as a result; thus, these genomes associated with the expression and/or infection by MSRV-1 are an integral part of the potential pathogenicity of this virus and therefore constitute diagnostic detection supports and particular therapeutic targets. Likewise, any agent which is associated with, or which is a cofactor for these interactions responsible for the pathogenicity in question, such as MSRV-2 or the gliotoxic factor described in the patent application published under the No. FR-2,716,198, can participate in the development of an overall and very effective strategy for therapeutic diagnosis, prognosis, monitoring and/or integrated therapy for MS in particular, but also for any other disease associated with the same agents.

In this context, a parallel discovery has been made in another autoimmune disease, rheumatoid arthritis (RA), which has been described in the French patent

application filed under the No. 95 02960. This discovery shows that, by applying methodological approaches similar to those which were used in the studies by the applicant on MS, it has been possible to identify a retrovirus
 5 expressed in RA which shares the sequences described for MSRV-1 in MS and also the coexistence of an MSRV-2-associated sequence which is also described in MS. As regards MSRV-1, the sequences commonly detected in MS and RA relate to the pol and gag genes. On the basis of
 10 current knowledge, it is possible to combine the gag and pol sequences described with the MSRV-1 strains expressed in these two diseases.

The present patent application has as its object various results, supplementary in relation to those
 15 already protected by the French patent applications:

No. 92/04322 of 03.04.1992, published under No. 2,689,519;

No. 92/13447 of 03.11.1992, published under No. 2,689,521;

20 No. 92/13443 of 03.11.1992, published under No. 2,689,520;

No. 94/01529 of 04.02.1994, published under No. 2,715,936;

25 No. 94/01531 of 04.02.1994, published under No. 2,715,939;

No. 94/01530 of 04.02.1994, published under No. 2,715,936;

No. 94/01532 of 04.02.1994, published under No. 2,715,937;

30 No. 94/14322 of 24.11.1994, published under No. 2,727,428;

No. 94/15810 of 23.12.1994, published under No. 2,728,585;

and

35 Patent Application WO 97/06260.

The present invention relates, first of all, to a nucleic material, which may consist of a retroviral material, in isolated or purified state, which may be understood or characterized in various ways:

5 - it comprises a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i);
10 and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively;

15 - it encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121,
20 SEQ ID NO: 135 and SEQ ID NO: 137;

 - its pol gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 112, SEQ ID NO: 124 and their complementary sequences;

25 - the 5' end of its pol gene starts at nucleotide 1419 of SEQ ID NO: 130;

 - its pol gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the
30 peptide sequence SEQ ID NO: 113;

 - the 3' end of its gag gene ends at nucleotide 1418 of SEQ ID NO: 130;

 - its env gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the
35 group which consists of SEQ ID NO: 117, and its complementary sequences;

- its env gene comprises a nucleotide sequence which starts at nucleotide 1 of SEQ ID NO: 117 and ends at nucleotide at nucleotide [sic] 233 of SEQ ID NO: 114;

5 - its env gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118;

10 - the U3R region of its 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of SEQ ID NO: 114;

- the RU5 region of its 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142;

15 - a retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114;

20 - the retroviral nucleic material as defined above is in particular associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

The invention also relates to a nucleotide fragment which corresponds to at least one of the following definitions:

25 - it comprises or consists of a nucleotide sequence chosen from the group which consists of (i)

the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences
30 complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively;

35 - it comprises or consists of a nucleotide sequence encoding a polypeptide having, for every

contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, 5 SEQ ID NO: 135 and SEQ ID NO: 137.

Other subjects of the present invention are the following:

- a nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or 10 rheumatoid arthritis, capable of hybridizing specifically with any fragment defined above and belonging to the genome of said retrovirus; it advantageously possesses from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides;

15 - a primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, which comprises a nucleotide sequence identical or equivalent to at least a portion of the nucleotide 20 sequence of a fragment defined above, in particular a nucleotide sequence having, for every series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment; preferably the nucleotide sequence of a primer of the 25 invention is chosen from SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and SEQ ID NO: 133;

- an RNA or a DNA, and in particular a replication 30 and/or expression vector, comprising a genomic fragment of the nucleic material or a fragment defined above;

- a peptide encoded by any open reading frame belonging to a nucleotide fragment defined above, in particular a polypeptide, for example oligopeptide forming 35 or comprising an antigenic determinant recognized by sera of patients infected with the MSRV-1 virus, and/or in whom

the MSRV-1 virus has been reactivated; a preferential peptide comprises a sequence identical, partially or completely, or equivalent to a sequence chosen from SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- a diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment defined above;

- a method for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, comprising the steps consisting of bringing an RNA and/or a DNA assumed to belong to or obtained from said retrovirus, or their complementary RNA and/or DNA, into contact with a composition comprising a nucleotide fragment defined above.

Before detailing the invention, various terms used in the description and the claims are now defined:

- strain or isolate is understood to mean any infectious and/or pathogenic biological fraction containing, for example, viruses and/or bacteria and/or parasites, generating a pathogenic and/or antigenic power, harbored by a culture or a live host; by way of example, a viral strain according to the preceding definition may contain a co-infectious agent, for example a pathogenic protist,

- the term "MSRV" used in the present description designates any pathogenic and/or infectious agent, as associated with MS, in particular a viral species, the attenuated strains of said viral species, or the interfering defective particles or particles containing co-encapsidated genomes or alternatively genomes recombined with a portion of the MSRV-1 genome, which are derived from this species. It is known that viruses and particularly viruses containing RNA exhibit variability,

following in particular relatively high rates of spontaneous mutation, which will be taken into account below to define the concept of equivalence,

5 - human virus is understood to mean a virus capable of infecting or of being harbored by human beings,

 - given all the natural or induced variations and/or recombination which may be encountered in practice in the present invention, the objects thereof, defined above and in the claims, have been expressed by comprising
10 the equivalents or derivatives of the various biological materials defined below, in particular homologous nucleotide or peptide sequences,

 - the variant of a virus or of a pathogenic and/or infectious agent according to the invention comprises at
15 least one antigen recognized by at least one antibody directed against at least one corresponding antigen of said virus and/or of said pathogenic and/or infectious agent, and/or a genome in which any portion is detected by at least one hybridization probe, and/or at least one
20 nucleotide amplification primer specific for said virus and/or pathogenic and/or infectious agent, under defined hybridization conditions well known to persons skilled in the art,

 - according to the invention, a nucleotide
25 fragment or an oligonucleotide or a polynucleotide is a stretch of monomers, or a biopolymer, characterized by the informational sequence of the natural nucleic acids, which is capable of hybridizing to any other nucleotide fragment under predefined conditions, it being possible for the
30 stretch to contain monomers of different chemical structures and to be obtained from a natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be identical to a genomic fragment of the MSRV-1 virus considered by
35 the present invention, in particular a gene of the latter, for example pol or env in the case of said virus;

- thus, a monomer may be a natural nucleic acid nucleotide in which the constituent components are a sugar, a phosphate group and a nitrogen base; in RNA, the sugar is ribose; in DNA, the sugar is 2-deoxyribose; depending on whether DNA or RNA is involved, the nitrogen base is chosen from adenine, guanine, uracil, cytosine, thymine; or the nucleotide may be modified in at least one of the three constituent components; by way of example, the modification may occur at the level of the bases, generating modified bases such as inosine, 5-methyl-deoxycytidine, deoxyuridine, 5-dimethylamine deoxyuridine [sic], 2,6-diamineopurine [sic], 5-bromodeoxyuridine and any other modified base promoting hybridization; at the level of the sugar, the modification may consist in the replacement of at least one deoxyribose with a polyamide, and at the level of the phosphate group, the modification may consist in its replacement with esters, in particular chosen from the esters of diphosphate, of alkyl and arylphosphonate and of phosphorothioate,

- "informational sequence" is understood to mean any ordered series of monomers, whose chemical nature and in which the order in a reference direction, constitute or otherwise a functional information of the same quality as that for the natural nucleic acids,

- hybridization is understood to mean the process during which, under appropriate operating conditions, two nucleotide fragments, having sufficiently complementary sequences, become annealed to form a complex, in particular a double or triple, structure, preferably in helical form,

- a probe comprises a nucleotide fragment synthesized by the chemical route or obtained by digestion or enzymatic cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 100 monomers, preferably 10 to 30 monomers, and

possessing a hybridization specificity under defined conditions; preferably, a probe possessing less than 10 monomers is not used alone, but is used in the presence of other probes which are equally short in length or otherwise; under certain specific conditions, it may be useful to use probes which are greater than 100 monomers in size; a probe may be used in particular for diagnostic purposes, and it may be, for example, capture and/or detection probes,

10 - the capture probe may be immobilized on a solid support by any appropriate means, that is to say directly or indirectly, for example by covalent bonding or passive adsorption,

 - the detection probe may be labeled by means of a
15 marker chosen in particular from radioactive isotopes, enzymes chosen in particular from peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorogenic or luminescent substrate, chromophoric chemical compounds, chromogenic, fluorogenic
20 or luminescent compounds, analogs of nucleotide bases, and biotin,

 - the probes used for diagnostic purposes of the invention may be used in all known hybridization techniques, and in particular the so-called "DOT-BLOT" technique, "SOUTHERN BLOT" technique, "NORTHERN BLOT"
25 technique which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, the SANDWICH technique; advantageously, the SANDWICH technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, it being
30 understood that the capture probe and the detection probe must have a nucleotide sequence which is at least partially different,

 - any probe according to the present invention may
35 hybridize in vivo or in vitro with the RNA and/or with the DNA, in order to block the replication, in particular

translation and/or transcription, phenomena and/or to degrade said DNA and/or RNA,

- a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, possessing hybridization specificity under defined conditions, for the initiation of an enzymatic polymerization, for example in an amplification technique such as PCR (Polymerase Chain Reaction), in an extension method such as sequencing, in a reverse transcription method and the like,

- two nucleotide or peptide sequences are said to be equivalent or derived with respect to each other, or with respect to a reference sequence, if functionally the corresponding biopolymers can play substantially the same role, without being identical, in relation to the application or use considered, or in the technique in which they are involved; particularly equivalent are two sequences obtained because of the natural variability, in particular spontaneous mutation, of the species from which they were identified, or induced mutation, as well as two homologous sequences, the homology being defined below,

- "variability" is understood to mean any spontaneous or induced modification of a sequence, in particular by substitution, and/or insertion, and/or deletion of nucleotides and/or of nucleotide fragments, and/or extension and/or shortening of the sequence at least at one of the ends; a nonnatural variability may result from the genetic engineering techniques used, for example from the choice of the degenerate or nondegenerate synthetic primers selected to amplify a nucleic acid; this variability may result in modifications of any starting sequence, considered as a reference, and which may be expressed by a degree of homology with respect to said reference sequence,

- homology characterizes the degree of identity of two compared nucleotide or peptide fragments; it is measured by the percentage identity which is in particular determined by direct comparison of nucleotide or peptide sequences, with respect to reference nucleotide or peptide sequences,

- any nucleotide fragment is said to be equivalent to or derived from a reference fragment if it has a nucleotide sequence equivalent to the sequence of the reference fragment; according to the preceding definition, in particular equivalent to a reference nucleotide fragment are:

(a) any fragment capable of hybridizing, at least partially, with the complementary to the reference fragment,

(b) any fragment whose alignment with the reference fragment leads to the identification of identical contiguous bases, in a greater number than with any other fragment obtained from another taxonomic group,

(c) any fragment resulting or capable of resulting from the natural variability of the species from which it is obtained,

(d) any fragment which may result from genetic engineering techniques applied to the reference fragment,

(e) any fragment, containing at least eight contiguous nucleotides, encoding a peptide homologous or identical to the peptide encoded by the reference fragment,

(f) any fragment different from the reference fragment through insertion, deletion, substitution of at least one monomer, extension, or shortening at least at one of its ends; for example, any fragment corresponding to the reference fragment, flanked at least at one of its ends by a nucleotide sequence not encoding a polypeptide,

- polypeptide is understood to mean in particular any peptide of at least two amino acids, in particular

oligopeptide, protein, extracted, separated, or substantially isolated or synthesized, through the involvement of humans, in particular those obtained by chemical synthesis, or through expression in a recombinant organism,

- polypeptide partially encoded by a nucleotide fragment is understood to mean a polypeptide having at least three amino acids encoded by at least nine contiguous monomers included in said nucleotide fragment,

- an amino acid is said to be analogous to another amino acid when their respective physicochemical characteristics, such as polarity, hydrophobicity and/or basicity, and/or acidity, and/or neutrality, are substantially the same; thus, a leucine is analogous to an isoleucine,

- any polypeptide is said to be equivalent to or derived from a reference polypeptide if the polypeptides compared have substantially the same properties, and in particular the same antigenic, immunological, enzymatic and/or molecular recognition properties; in particular equivalent to a reference polypeptide is:

(a) any polypeptide possessing a sequence in which at least one amino acid has been replaced by an analogous amino acid,

(b) any polypeptide having an equivalent peptide sequence, obtained by natural or induced variation of said reference polypeptide, and/or of the nucleotide fragment encoding said polypeptide,

(c) a mimotope of said reference polypeptide,

(d) any polypeptide from whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

(e) any polypeptide into whose sequence a modification of the side chains of the amino acids has been introduced, such as for example an acetylation of the

amine-containing functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

(f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as for example the
5 carba, retro, inverso, retro-inverso, reduced, and methylene-oxy bonds,

(g) any polypeptide in which at least one antigen is recognized by an antibody directed against a reference polypeptide,

10 - the percentage identity characterizing the homology between two peptide fragments compared is according to the present invention at least 50% and preferably at least 70%.

Given that a virus possessing a reverse
15 transcriptase enzymatic activity may be genetically characterized both in RNA and DNA form, both the viral DNA and RNA will be mentioned in order to characterize the sequences relative to a virus possessing such a reverse transcriptase activity, termed MSRV-1 according to the
20 present description.

The expressions of order which are used in the present description and the claims, such as "first nucleotide sequence", are not selected to express a particular order, but to define the invention more
25 clearly.

Detection of a substance or agent is understood below to mean an identification, a quantification or a separation or isolation of said substance or of said agent.

30 The invention will be understood more clearly on reading the detailed description which follows which is made with reference to the appended figures in which:

Figure 1 represents the general structure of the proviral DNA and the genomic RNA of MSRV-1.

35 Figure 2 represents the nucleotide sequence of the clone called CL6-5' (SEQ ID NO: 112) and three potential

reading frames in amino acids presented under the nucleotide sequence.

Figure 3 represents the nucleotide sequence of the clone called CL6-3' (SEQ ID NO: 114) and three potential
5 reading frames in amino acids presented under the nucleotide sequence.

Figure 4 represents the nucleotide sequence of the clone called C15 (SEQ ID NO: 117) and three potential
10 reading frames in amino acids presented under the nucleotide sequence.

Figure 5 represents the nucleotide sequence of the clone called 5M6 (SEQ ID NO: 120) and three potential
reading frames in amino acids presented under the nucleotide sequence.

15 Figure 6 represents the nucleotide sequence of the clone called CL2 (SEQ ID NO: 130) and three potential
reading frames in amino acids presented under the nucleotide sequence.

Figure 7 represents three potential reading frames
20 in amino acids expressed by pET28C-clone 2 and presented under the nucleotide sequence.

Figure 8 represents three potential reading frames in amino acids expressed by pET21C-clone 2 and presented
under the nucleotide sequence.

25 Figure 9 represents the nucleotide sequence of the clone called LB13 (SEQ ID NO: 141) and three potential
reading frames in amino acids presented under the nucleotide sequence.

Figure 10 represents the nucleotide sequence of
30 the clone called LA15 (SEQ ID NO: 142) and three potential
reading frames in amino acids presented under the nucleotide sequence.

Figure 11 represents the nucleotide sequence of the clone called LB16 (SEQ ID NO: 124) and three potential
35 reading frames in amino acids presented under the nucleotide sequence.

EXAMPLE 1: PREPARATION OF A CL6-5' REGION ENCODING THE N-TERMINAL END OF INTEGRASE AND OF A CL6-3' REGION CONTAINING THE 3' TERMINAL SEQUENCE OF THE MSRV-1 GENOME

5 A 3' RACE was carried out on the total RNA extracted from plasma from a patient suffering from MS. A healthy control plasma, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with an oligo dT primer identified by SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT TTT TTT TTT TTT TTT T 3') and the reverse transcriptase "ExpandTM RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the enzyme Klentaq (Clontech) under the following conditions: 94°C 5 min then 15 93°C 1 min, 58°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min, with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69
5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';
- 20 - 3' primer, identified by SEQ ID NO: 68

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR, 25 using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70
5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';
- 30 - 3' primer, identified by SEQ ID NO: 68

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1.

An amplification product of 1.9 Kb was obtained for the plasma of the MS patient. The corresponding 35 fragment was not observed for the healthy control plasma.

This amplification product was cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The 2 µl of DNA solution
5 were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCR[™] VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 12°C. The next steps were carried out in accordance with the
10 instructions for the TA Cloning kite (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The
15 plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert
20 after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[°] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119)
25 and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The clone obtained contains a CL6-5' region
30 encoding the N-terminal end of integrase and a CL6-3' region corresponding to the 3' terminal region of MSRV-1 and making it possible to define the end of the envelope (234 bp) and the U3 and R (401 bp) regions of the MSRV1 retrovirus.

35 The region corresponding to the N-terminal end of integrase is represented by its nucleotide sequence (SEQ

ID NO: 112) in Figure 27. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence, and the amineo [sic] acid sequence of the N-terminal end of integrase is identified
 5 by SEQ ID NO: 113.

The C16-3' region is represented by its nucleotide sequence (SEQ ID NO: 114) in Figure 3. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence. An amineo [sic]
 10 acid sequence corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 115.

In order to evaluate the promoter activity of the LTR obtained from clone 6 (cl6), a test of promoter activity using the enzyme CAT (chloramphenicol acetyl
 15 transferase) was carried out with the corresponding U3R region. In parallel, a clone containing the same U3R region of endogenous retroviral RNA expressed in normal placenta (PH74) and a clone (5M6) obtained from DNA were tested. The result presented in Figure 12 shows a very
 20 high promoter activity of the LTR derived from MS plasma (cl6) and a significantly much lower activity with the sequences of non-MS endogenous origin.

EXAMPLE 2: PREPARATION OF THE C15 CLONE CONTAINING
 25 THE REGION ENCODING A PORTION OF THE MSRV-1 RETROVIRUS ENVELOPE

A RT-PCR was carried out on the total RNA extracted from virions concentrated by ultracentrifugation of a synoviocyte culture supernatant obtained from an MS
 30 patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand™ RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the Expand™ Long Template PCR System (Boehringer) under the following
 35 conditions: 94°C 5 min then 93°C 1 min, 60°C 1 min, 68°C 3

min over 40 cycles and 68°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69
- 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';
- 3' primer, identified by SEQ ID NO: 116
- 5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR (except that 30 cycles were used instead of 40), using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70
- 5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';
- 3' primer, identified by SEQ ID NO: 116

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1. The primer SEQ ID NO: 116 is specific for the sequence FBd13 (also called B13) and is located in the conserved env region among the oncoretroviruses.

An amplification product of 1932 bp was obtained and cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after

staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the SP6 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior
 5 to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaqR FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A
 10 and 377 apparatus, according to the manufacturer's instructions.

The C15 clone obtained contains a region corresponding to the region of the MSRV-1 envelope of 1481 bp.

15 The env region of the C15 clone is represented by its nucleotide sequence (SEQ ID NO: 117) in Figure 5. The three potential reading frames of this clone are presented by their amino acid sequence under the nucleotide sequence. The reading frame corresponding to an MSRV-1
 20 structural env protein is identified by SEQ ID NO: 118.

From the defined sequences obtained from clones c16 and C15, it was possible to produce a plasmid construct encoding a complete envelope followed by the 3' LTR, as presented in Figure 13 with the corresponding
 25 reading frame.

EXAMPLE 3: PREPARATION OF A 5M6 CLONE CONTAINING THE SEQUENCES OF THE 3' TERMINAL REGION OF THE ENVELOPE, FOLLOWED BY THE MSRV-1 PROVIRAL TYPE U3, R AND U5
 30 SEQUENCES

A monodirectional PCR was carried out on the DNA extracted from immortalized B lymphocytes in culture from an MS patient. The PCR was carried out with Expand™ Long Template PCR System (Boehringer) under the following
 35 conditions: 94°C 3 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 10 cycles, then 93°C 1 min, 60°C 1 min with 15

sec of extension at each cycle, 68°C 3 min over 35 cycles and 68°C for 7 min and with a final reaction volume of 50 µl.

The primer used for the PCR identified by SEQ ID NO: 119 is 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT AAC CG 3';

The primers [sic] SEQ ID NO: 119 is specific for the env region of the C15 clone.

An amplification product of 1673 by was obtained and cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISMTM Ready Reaction AmpliTaq° FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The 5M6 clone obtained contains a region corresponding to the 3' region of the MSRV-1 envelope of 492 by followed by the regions U3, R and U5 (837 bp) of MSRV1.

The 5M6 clone is represented by its nucleotide sequence (SEQ ID NO: 120) in Figure 5. The three potential reading frames of this clone are presented by their amineo [sic] acid sequence under the nucleotide sequence. The reading frame corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 121.

EXAMPLE 4: PREPARATION OF THE LB16 CLONE CONTAINING THE REGION ENCODING THE MSRV-1 RETROVIRUS INTEGRASE

An RT-PCR was carried out on the total RNA treated with DNaseI and extracted from a choroid plexus obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand TM RT" from Boehringer according to the conditions recommended by the company. A "no RT" control was carried out in' parallel on the same material. A PCR was carried out with Taq polymerase (Perkin Elmer) under the following conditions: 95°C 5 min, then 95°C 1 min, 55°C 1 min, 72°C 2 min over 35 cycles and 72°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 122
5' GGC ATT GAT AGC ACC CAT CAG 3';
- 3' primer, identified by SEQ ID NO: 123
5' CAT GTC ACC AGG GTG GAA TAG 3'

The primer SEQ ID NO: 122 is specific for the pol region of MSRV-1 and more precisely similar to the integrase region described above. The primer SEQ ID NO 123 was defined on sequences of the clones obtained during preliminary tests.

An amplification product of about 760 bp was obtained only in the test with RT and was cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were

carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaqR FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB16 clone obtained contains the sequences corresponding to integrase. The nucleotide sequence of this clone was identified by SEQ ID NO: 124 in Figure 11, three reading frames are determined.

25

EXAMPLE 5: PREPARATION OF A CLONE 2, CL2, CONTAINING IN 3' A PORTION HOMOLOGOUS TO THE POL GENE, CORRESPONDING TO THE PROTEASE GENE, AND TO THE GAG GENE (GM3) CORRESPONDING TO THE NUCLEOCAPSID, AND A NEW 5' CODING REGION, CORRESPONDING TO THE GAG GENE MORE SPECIFICALLY THE TEMPLATE AND THE CAPSID of MSRV-1.

A PCR amplification was carried out on the total RNA extracted from 100 µl of plasma from a patient suffering from MS. A water control, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with 300 pmol of a random primer

35

(GIBCO-BRL, France) and the reverse transcriptase "Expand RT" (BOEHRINGER MANNHEIM, France) according to the conditions recommended by the company. An amplification by PCR ("polymerase chain reaction") was carried out with the enzyme Taq polymerase (Perkin Elmer, France) using 10 µl of cDNA under the following conditions: 94°C 2 min, 55°C 1 min, 72°C 2 min then 94°C 1 min, 55°C 1 min, 72°C 2 min over 30 cycles and 72°C for 7 min with a final reaction volume of 50 µl.

10 Primers used for the PCR amplification:
 - 5' primer, identified by SEQ ID NO: 126
 5' CGG ACA TCC AAA GTG ATG GGA AAC G 3';
 - 3' primer, identified by SEQ ID NO: 127
 5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

15 A second amplification by so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR, using 10 µl of the amplification product derived from the first PCR.

20 Primers used for the amplification by seminested PCR:

 - 5' primer, identified by SEQ ID NO: 128
 5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3';
 25 3' primer, identified by SEQ ID NO: 129
 5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

 The primers SEQ ID NO: [lacuna] and SEQ ID NO: [lacuna] are specific for the pol region, clone G+E+A, more specifically the E region: nucleotide position No. 30 423 to No. 448. The primers used in the 5' region were defined on sequences of clones obtained during preliminary tests.

 An amplification product of 1511 bp was obtained from the RNA extracted from the plasma of an MS patient.
 35 The corresponding fragment was not observed for the water

control. This amplification product was cloned in the following manner.

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kitTM. The 2 μ l of DNA solution were mixed with 5 μ l of sterile distilled water, 1 μ l of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 μ l of "pCRTM VECTOR" (25 ng/ml) and 1 μ l of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit[®] (Invitrogen). The mixture was plated after transformation of the ligation into E. coli INV α F' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme EcoRI and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISMTM Ready Reaction Amplitaq[®] FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The clone obtained, called CL2, contains a C-terminal region similar to the 5' terminal region of the clones G+E+A of MSRV-1, which makes it possible to define the C-terminal region of the gag gene and a new region corresponding to the N-terminal region of the MSRV-1 gag gene.

CL2 makes it possible to define a region of 1511 by having an open reading frame in the N-terminal region of 1077 bp encoding 359 amino acids and a nonopen reading frame of 454 bp corresponding to the C-terminal region of the MSRV-1 gag gene.

The nucleotide sequence of CL2 is identified by SEQ ID NO: 130. It is represented in Figure 6 with the potential reading frames in amineo [sic] acid.

The 1077 bp fragment of CL2 encoding 359 amino acids was amplified by PCR with the Pwo enzyme (5U/ μ l) (Boehringer Mannheim, France) using 1 μ l of the DNA minipreparation of clone 2 under the following conditions: 95°C 1 min, 60°C 1 min, 72°C 2 min over 25 cycles and with a final reaction volume of 50 μ l with the aid of the primers:

5' primer (*Bam*HI), identified by SEQ ID NO: 132

5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer), and

3' primer (*Hind*III), identified by SEQ ID NO: 133

5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer) corresponding, respectively, to the nucleotide sequence of clone 2 at position -9 to 21 and 1066 to 1095.

The fragment obtained by PCR was linearized with *Bam*HI and *Hind*III and subcloned into the expression vectors pET28C and pET21C (NOVAGEN) linearized with *Bam*HI and *Hind*III. The sequencing of the DNA of the 1077 bp fragment of clone 2 in the two expression vectors was carried out according to the method recommended for the use of the sequencing kit "PRISM™ Ready Reaction Amplitae® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref.402119 and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The expression of the nucleotide sequence of the 1077 bp fragment of clone 2 by the expression vectors

pET28C and pET21C are identified by SEQ ID NO: 135 and SEQ ID NO: 137, respectively.

EXAMPLE 6: EXPRESSION OF CLONE 2 IN *ESCHERICHIA*

5 *COLI*

The constructs pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) synthesize, in the bacterial strain BL21 (DE3), a protein fused at the N- and Cterminus for the vector pET28C and the C-terminus for the vector
10 pET21C with 6 Histidines, having an apparent molecular mass of about 45 kDa, identified by SDS-PAGE polyacrylamide gel electrophoresis (SDS = Sodium Dodecyl Sulfate) (Laemmli, 1970 (1)). The reactivity of the protein was demonstrated towards an anti-Histidine
15 monoclonal antibody (DIANOVA) by the Western-blot technique (Towbin et al., 1979 (2)).

The recombinant proteins pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were visualized by SDS-PAGE in the insoluble fraction after enzymatic digestion of the
20 bacterial extracts with 50 μ l of lysozyme (10 mg/ml) and ultrasound lysis.

The antigenic properties of the recombinant antigens pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were tested by Western blotting () [sic] after
25 solubilization of the bacterial pellet with 2% SDS and 50 mM β -mercaptoethanol. After incubation with sera from patients suffering from multiple sclerosis, the sera from neurological controls and the sera from controls at the Blood Transfusion Center (CTS), the immunocomplexes were
30 detected with the aid of an alkaline phosphatase-coupled goat serum anti-human IgG and anti-human IgM.

The results are presented in the table below.

TABLE

Reactivity of sera affected by multiple sclerosis and
35 controls with the MSRV-1 recombinant protein gag clone 2

(1077 bp) = pET21C-clone 2 (1077 bp) and pET28C-clone 2 (1077 bp)^a

DISEASE	NUMBER OF INDIVIDUALS TESTED	NUMBER OF POSITIVE INDIVIDUALS
MS	15	6 2 (+++), 2 (++), (2 (+)
NEUROLOGICAL		
CONTROLS	2	1(+++)
HEALTHY		
CONTROLS (CTS)	22	1(+/-)

(a) The strips containing 1.5 µg of recombinant antigen pET-gag clone 2 (1077 bp) exhibit reactivity against sera diluted 1/100. The Western-Blot interpretation is based on the presence or absence of a specific pET-gag clone 2 (1077 bp) band on the strips. Positive and negative controls are included in each experiment.

These results show that, under the technical conditions used, about 40% of the human sera affected by multiple sclerosis which were tested react with the recombinant proteins pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp). Reactivity was observed on a neurological control and it is of interest to note that the RNAs extracted from this serum, after the reverse transcriptase step, are also amplified by PCR in the pol region. This suggests that people who have not declared MS may also harbor and express this virus. On the other hand, an apparently healthy control (CTS donor) possesses anti-gag (clone 2, 1077 bp) antibodies. This is compatible with an immunity acquired against MSRV-1 independently of a declared associated autoimmune disease.

EXAMPLE 7: PREPARATION OF AN LB13 CLONE CONTAINING IN 3' A PORTION HOMOLOGOUS TO CLONE 2 CORRESPONDING TO THE GAG

GENE AND IN 5' A PORTION HOMOLOGOUS TO THE 5M6 CLONE
CORRESPONDING TO THE U5 LTR REGION

An RT-PCR ("reverse transcriptase-polymerase chain
reaction") was carried out using total RNA extracted from
5 virions, obtained from supernatants of B lymphocyte cells
of patients suffering from multiple sclerosis,
concentrated by ultracentrifugations. The synthesis of
cDNA was carried out with a specific primer SEQ No. XXX
and the reverse transcriptase "Expand™ RT" from BOEHRINGER
10 MANNHEIM according to the conditions recommended by the
company.

Primer used for the synthesis of the cDNA, identified by
SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

15 A PCR amplification was carried out with Taq
polymerase (Perkin Elmer, France) under the following
conditions: 94°C 1 min, 55°C 1 min, 72°C 2 min over 35
cycles at 72°C for 7 min and with a final reaction volume
of 100 µl.

20 Primers used for the PCR amplification:

- 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- 3' primer, identified by SEQ ID NO: 138

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

25 A second so-called "seminested" PCR amplification
was carried out with a 3' primer situated inside the
region already amplified. This second amplification was
carried out under the same experimental conditions as
those used during the first amplification, using 10 µl of
30 the amplification product derived from the first PCR.

Primers used for the "seminested" PCR amplification:

- 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- 3' primer, identified by SEQ ID NO: 140

35 5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

The primers SEQ ID NO: 138 and SEQ ID NO: 140 are specific for the gag region, clone 2 nucleotide position No. 373-397 and No. 433-456. The primers used in the 5' region were defined on sequences of the clones obtained during preliminary tests.

An amplification product of 764 bp was obtained and cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kitTM. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCRTM VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit[®] (Invitrogen). The mixture was plated after transformation of the ligation into E. coli INVαF' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme EcoRI and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISMTM Ready Reaction Amplitaq[®] FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB13 clone obtained contains an N-terminal region of MSRV-1 gag gene homologous to clone 2 and an LTR corresponding to a portion of the U5 region. Between the U5 region and gag, a binding site for the transfer RNAs, the PBS "primer binding site", was identified.

The nucleotide sequence of the 764 bp fragment of the LB13 clone in the plasmid "pCRTM vector" is represented in the identifier SEQ ID NO: 141.

The binding site for the transfer RNAs, having a sequence of PBS tryptophan type, was identified at nucleotide position No. 342-359 of the LB13 clone.

As this same PBS was found in the endogenous copies homologous to MSRV1, the endogenous family thus defined is henceforth called HERV W, according to the nomenclature proposed for the endogenous retrovirus families (W=tryptophan).

A short ORF of about 65 amino acids was found in the U5 region of the 5' LTR of the LB13 clone.

Sequence of the ORF:

PMASNRAITLTAWSKIPFLGIRETKNPRSENTRLATMLEAAHHHFGSSPPLSWEL
WEQGPQVTIW.

The corresponding nucleotide sequence starting at an ATG codon is capable of being expressed in a subgenomic DNA from a proviral LTR (U3RU5).

Another clone, called LA15, was obtained on the total RNA extracted from virions concentrated by ultracentrifugation from a culture supernatant of synoviocytes obtained from a patient suffering from rheumatoid arthritis. The strategy for amplifying and cloning the LA15 clone is exactly the same which was used for the LB13 clone.

The nucleotide sequence of the LA15 clone, which is represented in the identifier SEQ ID NO: 142, is very similar to the LB13 clone. This suggests that the MSRV-1 retrovirus detected in multiple sclerosis has sequences which are similar to those found in rheumatoid arthritis.

BIBLIOGRAPHIC REFERENCES

- (1) Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4.
5 Nature. (1970). 227: 680-685.
- (2) Towbin H., Staehelin T. & Gordon J. Electrophoretic transfer of proteins from polyacryalmide gels to nitrocellulose sheets: procedure and some
10 applications. Proc. Natl. Acad. Sci. USA.
(1979). 76: 4350-4354.

CLAIMS

1. Nucleic material, in isolated or purified state, comprising a nucleotide sequence chosen from the group
5 which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in
10 particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.
2. Nucleic material, in isolated or purified state,
15 encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and
20 SEQ ID NO: 137.
3. Retroviral nucleic material, whose pol gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 112, SEQ ID NO: 124 and their complementary sequences.
- 25 4. Retroviral nucleic material, in which the 5' end of the pol gene starts at nucleotide 1419 of SEQ ID NO: 130.
5. Retroviral nucleic material, in which the pol gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably
30 at least 70% homology with the peptide sequence SEQ ID NO: 113.
6. Retroviral nucleic material, in which the 3' end of the gag gene ends at nucleotide 1418 of SEQ ID NO: 130.
7. Retroviral nucleic material, in which the env gene
35 comprises a nucleotide sequence identical or equivalent to

a sequence chosen from the group which consists of SEQ ID NO: 117, and its complementary sequences.

8. Retroviral nucleic material, in which the env gene comprises a nucleotide sequence which starts at nucleotide 1 of SEQ ID NO: 117 and ends at nucleotide at nucleotide [sic] 233 of SEQ ID NO: 114.
9. Retroviral nucleic material, in which the env gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118.
10. Retroviral nucleic material in which the U3R region of the 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of SEQ ID NO: 114.
11. Retroviral nucleic material in which the RU5 region of the 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142.
12. Retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114.
13. Retroviral nucleic material according to any one of the preceding claims, characterized in that it is associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.
14. Nucleotide fragment comprising a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.
15. Nucleotide fragment according to Claim 14, consisting of a nucleotide sequence chosen from the group which

consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.

16. Nucleotide fragment comprising a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

17. Nucleotide fragment according to claim 16, consisting of a nucleotidesequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

18. Nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, characterized in that it is capable of hybridizing specifically with any fragment according to any one of claims 14 to 17, belonging to the genome of said retrovirus.

19. Probe according to claim 18, characterized in that it possesses from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides.

20. Primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, characterized in that it comprises a nucleotide sequence identical or

equivalent to at least a portion of the nucleotide sequence of a fragment according to any one of claims 8 to 11, in particular a nucleotide sequence having, for every series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment.

21. Primer according to claim 20, characterized in that its nucleotide sequence is chosen from SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and SEQ ID NO: 133.

22. RNA or DNA, and in particular replication and/or expression vector, comprising a genomic fragment of the nucleic material according to any one of claims 1 to 7 or a fragment according to any one of claims 14 to 17.

23. Peptide encoded by any open reading frame belonging to a nucleotide fragment according to any one of claims 14 to 17, in particular a polypeptide, for example oligopeptide forming or comprising an antigenic determinant recognized by sera of patients infected with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated.

24. Peptide according to claim 23 comprising a sequence identical, partially or completely, or equivalent to a sequence chosen from SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

25. Diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment according to any one of claims 14 to 17.

26. Method for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, characterized in that an RNA and/or a DNA assumed to belong to or obtained from said retrovirus, or their complementary RNA and/or DNA, is brought into

contact with a composition comprising a nucleotide fragment according to any one of claims 14 to 17.

SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT

34

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GCCATCAAGC CACCCAAGAA CTCTTAAC TT

30

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCAATAGCCA GACCATTATA TACACTAATT

30

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 310 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```

GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC CCCAGTACTC   60
AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT   120
AGCCACTGAG GAAGGAAAAA TACTTTCACC TGCAGCTAAC CAACAGAAAT TACTTAAAC   180
CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT   240
TACTGGACCA GGCCTTTTCA AAACATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA   300
AAGAAATAAT                                     310

```

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

LIEGPLVWGNPLWETKPQYSAGKIEXETSQGHTFLPSRWLATEEGKILSPAANQQKLLKTLHQTFHLGID
 STHQMAKLLFTGPGLFKTIKKIVRGCEVCQRNN

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 635 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```

CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC TGTAAAACTA   60
CAAATTGTTC TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC   120
CTGGACCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG   180
GAAATCTCAA CTGCACAACC CCTACTATGC CCAATTTCAG CGGGAAGCAG TTAGAGCGGT   240
CATCAGCCAA CCTCCCCAAC AGCACTTGGG TTTTCCTGTT GAGAGGGGGG ACTGAGAGAC   300
AGGACTAGCT GGATTTCTTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT   360
GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCCGACCAAT CAGAGAGCTC   420
ACTAAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG   480
AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAAACCCA GGCATTCGAG CCGGCAACGG   540

```

CAACCCCCTT TGGGTCCCCT CCCTTTGTAT GGGCGCTCTG TTTTCACTCT ATTTCACTCT 600
 ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

PCIFNLLVKFVSSRIKTVKLQIVLQMEHQMESMTKIHGRPLDRPASPCSDVNDIEGTPPEEISTAQPLL
 C
 PNSAGSS

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

TGGGGTTCCA TTTGTAAGAC CATCTGTAGC TT 32

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1481 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60
 GCACCCCTC CATGCTGCTG TACAACAGT AGCTCCCTT ACCAAGAGT TCTATGAAGA 120
 ACGCGGCTTC CTGGAAATAT TGATGCCCCA TCATATAGGA GTTTATCTAA GGGAACTCC 180

```

ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG 240
CATGCAAATA CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 300
GGAGCCACTG TCTGTTGGAC TTACTTCACC CATAACAGTA TGTCTGATGG GGGTGGGAATT 360
CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420
AGCACCCCTA GCCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480
CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540
CAAAACCCTA CTAAGTGTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTCAATC 600
CCTGTTCTTG AACAAATGGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 660
GGACCTCTTG TTTCCAATCT GGAAATAACC CATACTCAA ACCTCACCTG TGTAATAATT 720
AGCAATACTA TAGACACAAC CAGCTCCCAA TGCATCAGGT GGGTAACACC TCCCACACGA 780
ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG 840
AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC 900
ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960
CTTCCTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020
ACAACCTCTA CTCAGTTCTA CTACAAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080
GTCAGTACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 1140
CAAAATCGAA GAGCTTTAGA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTTTA 1200
GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260
ATTTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAAA ACACCGAACG CTGGGGCCTC 1320
CTCAGCCAAT GGATGCCCTG GGTCTCCCC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380
TTACTCCTCT TTGGACCCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440
GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

```

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 493 amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

```

MALPYHTFLFTVLLPPFALTAPPPCCCTSSSPYQEFLEXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPR
NCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTFHTSMSGGGIQGQAREKQVKEAISQLTRGH
STPSPYKGLVLSKLHETLRTHRLVSLFNNTTLRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNFS
TEINTTSVLVGPLVSNLEITHTSNLTVCVKFSNTIDTSSQCIRWVTPPTRIVCLPSGIFVCGTSAYHCL
NGSSESMCFLSFLVPPMTIYTEQDLYNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYK
SQTINGDMEQVTDLSLVTLDQLNSLAADVLLQNRRLDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE
IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMV
MEP

```

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG

32

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1329 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT GTTTATTTTT 60
 AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA 120
 AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT 180
 CCTCAGCCAA TGGATGCCCT GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT 240
 TTTACTCCTC TTTGGACCCT GTATCTTCAA CTTCCTTGTT AAGTTTGTCT CTTCCAGAAT 300
 TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT CCATGACTAA 360
 AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA 420
 AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG 480
 AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA 540
 GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT 600
 ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG 660
 ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA CAGGAGGTAA AGCAATAGCC 720
 AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA 780
 TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT 840
 CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG 900
 CTCAAGCTGA GCTTTTGTTT GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT 960
 GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT 1020
 ACCCATTGCC ACTCCCGATC AGGCTAAAGG CTTGCCATTG TTCCTGCATG GCTAAGTGCC 1080
 TGGGTTTGTC CTAATAGAAC TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC 1140
 CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGGTA 1200
 TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TGCCACCATT TGGGAAGTGG 1260

CCCACTGCCA TTTTGGTAGC GGCCCACCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320
CCAGTAACA 1329

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

QNRRALDLLTAKRGGTCLFLGEECCXYVNQSGIITEKVKEIXDRIXCRAEDLQNTAPWGLLSQWMPWTLP
FLGPLAAIIFLLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEPQMOSMTKIYRGPLDRPARLCSDVNDIE
VTPPEEISTAQPLLHSNSVGSS

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GGCATTGATA GCACCCATCA G

21

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

CATGTCACCA GGGTGGAATA G

21

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs

- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

(2) INFORMATION FOR SEQ ID NO: 126:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

CATGTCACCA GGGTGGAATA G

25

(2) INFORMATION FOR SEQ ID NO: 127:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

GGACAGGAAA GTAAGACTGA GAAGGC

26

(2) INFORMATION FOR SEQ ID NO: 128:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: base pairs
 - (B) TYPE: nucleotide
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

See Example 5

(2) INFORMATION FOR SEQ ID NO: 129:**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

See Example 5

(2) INFORMATION FOR SEQ ID NO: 130:**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 1511 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

```

CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG   60
ATTTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG  120
GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA  180
GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACCTC  240
ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA  300
CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC  360
GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCCCG  420
ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTCGGC CCAGCCAGAG TGCCTGTACC  480
TTTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAATTCT CAGATAACCC  540
TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA  600
TATAATGTTA CTACTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCG CTGTAACCTGC  660
AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA  720
GGAAAGAACA ACTCCACAG GCCAGCAGGC AGTTCCAGT GTAGACCCTC ATTGGGACAC  780
AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTTGCTA ACTTGCGTGC TAGAAGGACT  840
GAGGAAACT AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA  900
GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC  960
CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020
AGCTGCAGAC ATTAGAAAAA ACTTCAAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080
ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140
CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCCCTCAGGC 1200
AAGCAGACTT TGGAGGCTCT GCAAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260

```

CTGGCTTCCA GTGCGGTCTA CAAGGACACT TTAAAAAGA TTATCCAAGT AGAAATAAGC 1320
 CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380
 GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440
 GCCCCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTTGACCA 1500
 TTGAGAGCCA A 1511

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 347 amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

LERILENWDQCDTQTLRKKRFFFCSTAWPQYPLQGRETWLPEGSINYNIILQLDLFCRKEGKWSEVPYV
 QTFSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPTPSSTNKDPPLTQTVQKEIDKGVNNEPKSANIPR
 LCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLKGFSDNPDGYIDVLQGLGQSFDLTWRDIMLLLNQTLT
 PNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTGQAVPSVDPHWDTESEHGDWCHKHLLTCVLEGL
 RKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRKHTSLSPDSIEGQLILKDKFITQSAADIRKNFKS
 LP

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TGCTGGAATT CGGGATCCTA GAACGTATTC

30

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:
AGTTCTGCTC CGAAGCTTAG GCAGACTTTT

30

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRILERILENWDQCDTQTLRKKRFFFCSTAWPQYPLQGR
ETWLPEGSINYNIIQLDLFCRKEGKWSEVPYVQTFFSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPT
PSSTNKDPPLTQTVQKEIDKGVNNEPKSANIPRLCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLKGFSD
NPDGYIDVLQGLGQSFDLTWRDIMLLLNQTLTPNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTG
QQAVPSVDPHWDTESEHGDWCHKHLLTCVLEGLRKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRK
HTSLSPDSIEGQLILKDKFITQSAADIRKNFKSLPKLAAALEHHHHHH

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: amino acids
- (B) TYPE: peptide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

MASMTGGQQMGRILERILENWDQCDTQTLRKKRFFFCSTAWPQYPLQGRETWLPEGSINYNIIQLDLF
CRKEGKWSEVPYVQTFFSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPTPSSTNKDPPLTQTVQKEIDK
GVNNEPKSANIPRLCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLKGFSDNPDGYIDVLQGLGQSFDLT
WRDIMLLLNQTLTPNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTGQQAVPSVDPHWDTESEHGDW
CHKHLLTCVLEGLRKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRKHTSLSPDSIEGQLILKDKFI
TQSAADIRKNFKSLPKLAAALEHHHHHH

(2) INFORMATION FOR SEQ ID NO: 138:**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

CTTGGAGGGT GCATAACCAG GGAAT 25

(2) INFORMATION FOR SEQ ID NO: 139:**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

TGTCCGCTGT GCTCCTGATC 20

(2) INFORMATION FOR SEQ ID NO: 140:**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

CTATGTCCTT TTGACTGTT TGGGT 25

(2) INFORMATION FOR SEQ ID NO: 141:**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 764 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60
 CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTCATC CTAATCGAGC TGAACACTAG 120
 TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180

CTGCATGGTC	CAAGATTCCA	TTCCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	240
ACACAAGGCT	TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTTGGAAGC	AGCCCGCCAC	300
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	AGGTAACAAT	TTGGTGACCA	CGAAGGGACC	360
TGAATCCGCA	ACCATGAAGG	GATCTCCAAA	GCAATTGGAA	ATGTTCTCTC	CAAGGCAAAA	420
ATGCCCCTAA	GATGTATTCT	GGAGAATTGG	GACCAATTTG	ACCCTCAGAC	AGTAAGAAAA	480
AAATGACTTA	TATTCTTCTG	CAGTACCGCC	CTGGCCACGA	TATCCTCTTC	AAGGGGGAGA	540
AACCTGGCCT	CCTGAGGGAA	GTATAAATTA	TAACACCATC	TTACAGCTAG	ACCTGTTTTG	600
TAGAAAAGGA	GGCAAATGGA	GTGAAGTGCC	ATATTTACAA	ACTTTCTTTT	CATTAAAAGA	660
CAACTCGCAA	TTATGTTAAC	AGTGTGATTT	GTGTTCTTAC	ACGGAAGCCC	TCAGATTCTA	720
CTCCCCACCC	CCGGCATCTC	CCCTGAATCC	CTCCCCAACT	TATT		764

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 800 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

TGTCCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTGCC	TCTCCCAATT	GGGCTAAAGG	60
CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTCATC	CTAATCGAGC	TGAACACTAG	120
TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	CCATGGCTTC	TAATAGAGCT	ATAAACTCA	180
CTGCATGGTC	CAAGATTCCA	TTCCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	240
ACACAAGGCT	TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTTGGAAGC	GGCCCGCCAC	300
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	CAGGTAACAA	TTTGGTGACC	ACGAAGGGAC	360
CTGAATCCGC	AACCATGAAG	GGATCTCCAA	AGCAATTGGA	AATGTTCTTC	CCAAGGCAAA	420
AATGCCCCTA	AGATGTATTG	TGGAGAATTG	GGACCAATCT	GACCTCAGA	CAGTAAGAAA	480
AAAAATGACT	TATATTCTTC	TGCAGTACCG	CCTGGCCACG	GATATCCTCT	TCAAGGGGGA	540
GAAACCTGGC	CTCCTGAGGG	AAGTATAAAT	TATAACACCA	TCTTACAGCT	AGACCTGTTT	600
TGTAGAAAAG	GAGGCAAATG	GAGTGAAGTG	CCATATTTAC	AAACTTTCTT	TTCATTAAAA	660
GACAACTCGC	AATTATGTAA	ACAGTGTGAT	TTGTGTCCTA	CAGGAAGCCC	TCAGATCTAC	720
CTCCCTACCC	CGGCATCTCC	CTGACTCCTT	CCCCAACTAA	TAAGGACCCA	CTTCAGCCCA	780
AACAGTCCAA	AAGGACATAG					800

ABSTRACT

RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES

Nucleic material, in isolated or purified state, comprising a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively, and uses for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis.

FIG. 1

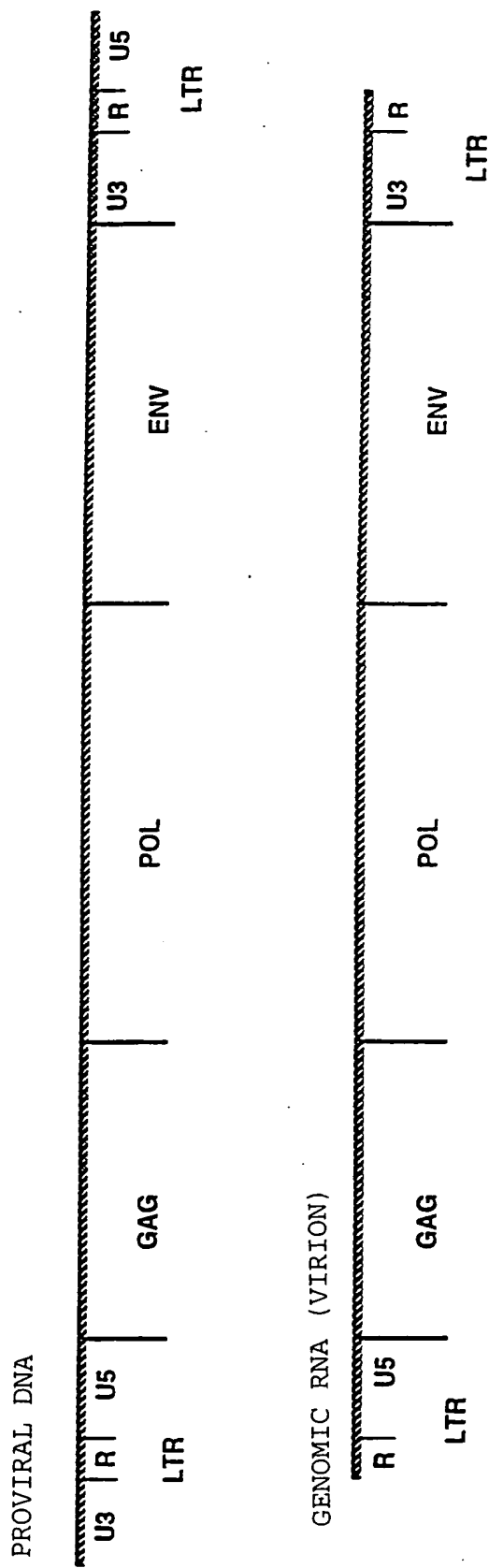


FIG 2

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCTTATAGAA	GGACCCCTAG	TATGGGGTAA	TCCCTCTGG	GAAACCAAGC	50
A Y R R	T P S	M G .	S P L G	N Q A	
L I E	G P L V	W G N	P L W	E T K P	
L . K	D P .	Y G V I	P S G	K P S	
CCAGTACTC	AGCAGGAAAA	ATAGAATAGG	AAACCTCACA	AGGACATACT	100
P V L	S R K N	R I G	N L T	R T Y F	
Q Y S	A G K	I E .	E T S Q	G H T	
P S T Q	Q E K	. N R	K P H K	D I L	
TTCTCCCTC	CCAGATGGCT	AGCCACTGAG	GAAGGAAAAA	TACTTTTACC	150
P P L	Q M A	S H .	G R K N	T F T	
F L P S	R W L	A T E	E G K I	L S P	
S S P	P D G .	P L R	K E K	Y F H L	
TGCAGCTAAC	CAACAGAAAT	TACTTAAAAC	CCTTCACCAA	ACCTTCCACT	200
C S .	P T E I	T . N	P S P N	L P L	
A A N	Q Q K L	L K T	L H Q	T F H L	
Q L T	N R N	Y L K P	F T K	P S T	
TAGGCATTGA	TAGCACCCAT	CAGATGGCCA	AATTATTATT	TACTGGACCA	250
R H .	. H P S	D G Q	I I I	Y W T R	
G I D	S T H	Q M A K	L L F	T G P	
. A L I	A P I	R W P	N Y Y L	L D Q	
GGCCTTTTCA	AAACTATCAA	GAAGATAGTC	AGGGGCTGTG	AAGTGTGCCA	300
P F Q	N Y Q	E D S Q	G L .	S V P	
G L F K	T I K	K I V	R G C E	V C Q	
A F S	K L S R	R . S	G A V	K C A K	
AAGAAATAAT					310
K K .					
R N N					
E I					

FIG 2 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCCTGTATCT	TTAACCTCCT	TGTTAAGTTT	GTCTCTTCCA	GAATCAAAC	50
P C I F	N L L	V K F	V S S R	I K T	
P V S	L T S L	L S L	S L P	E S K L	
L Y L	. P P	C . V C	L F Q	N Q N	
TGTAAACTA	CAAATGTTC	TTCAAATGGA	GCACCAGATG	GAGTCCATGA	100
V K L	Q I V L	Q M E	H Q M	E S M T	
. N Y	K L F	F K W S	T R W	S P .	
C K T T	N C S	S N G	A P D G	V H D	
CTAAGATCCA	CCGTGGACCC	CTGGACCGGC	CTGCTAGCCC	ATGCTCCGAT	150
K I H	R G P	L D R P	A S P	C S D	
L R S T	V D P	W T G	L L A H	A P M	
. D P	P W T P	G P A	C . P	M L R C	
GTTAATGACA	TTGAAGGCAC	CCCTCCCGAG	GAAATCTCAA	CTGCACAACC	200
V N D I	E G T	P P E	E I S T	A Q P	
L M T	L K A P	L P R	K S Q	L H N P	
. . H	. R H	P S R G	N L N	C T T	
CCTACTATGC	CCCAATTCAG	CGGGAAGCAG	TTAGAGCGGT	CATCAGCCAA	250
L L C	P N S A	G S S	. S G	H Q P T	
Y Y A	P I Q	R E A V	R A V	I S Q	
P T M P	Q F S	G K Q	L E R S	S A N	
CCTCCCCAAC	AGCACTTGGG	TTTTCTGT	GAGAGGGGG	ACTGAGAGAC	300
S P T	A L G	F S C	. E G G	L R D	
P P Q Q	H L G	F P V	E R G D	. E T	
L P N	S T W V	F L L	R G G	T E R Q	
AGGACTAGCT	GGATTTCCTA	GGCCAACGAA	GAATCCCTAA	GCCTAGCTGG	350
R T S W	I S .	A N E	E S L S	L A G	
G L A	G F P R	P T K	N P .	A . L G	
D . L	D F L	G Q R R	I P K	P S W	

FIG 3

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GAAGGTGACT	GCATCCACCT	CTAAACATGG	GGCTTGCAAC	TTAGCTCACA	400
K V T A S T S	K H G A C N	L A H T			
R . L H P P	L N M G	L A T .	L T		
E G D C I H L	. T W G L Q L	S S H			
CCCCACCAAT	CAGAGAGCTC	ACTAAAATGC	TAATTAGGCA	AAAATAGGAG	450
R P I R E L	T K M L	I R Q K .	E		
P D Q S E S S	L K C .	L G K N R R			
P T N Q R A H	. N A N .	A K I G G			
GTAAAGAAAT	AGCCAATCAT	CTATTGCCTG	AGAGCACAGC	GGGAGGGACA	500
V K K . P I I	Y C L R A Q R	E G Q			
. R N S Q S S	I A . E H S	G R D K			
K E I A N H	L L P E S T A	G G T			
AGGATCGGGA	TATAAACCCA	GGCATTGAG	COGGCAACGG	CAACCCCTT	550
G S G Y K P R	H S S R Q R	Q P P L			
D R D I N P	G I R A G N G	N P L			
R I G I . T Q	A F E P A T A	T P F			
TGGGTCCCTT	CCCTTTGAT	GGGCGCTCTG	TTTCACTCT	ATTCACTCT	600
G P L P L Y	G R S V F T L	F H S			
W V P S L C M	G A L F S L Y	F T L			
G S P P F V W	A L C F H S	I S L Y			
ATTAAATCTT	GCAACTGAAA	AAAAAAAAAA	AAAAA		635
I K S C N .	K K K K K				
L N L A T E	K K K K K				
. I L Q L K	K K K K K				

FIG 4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCTCC	CTTATCATAC	TTTTCTCTTT	ACTGTTCTCT	TACCCCTTT	50
M A L P	Y H T	F L F	T V L L	P P F	
W P S	L I I L	F S L	L F S	Y P L S	
G P P	L S Y	F S L Y	C S L	T P F	
CGCTCTCACT	GCACCCCTC	CATGCTGCTG	TACAACCAGT	AGCTCCCTT	100
A L T	A P P P	C C C	T T S	S S P Y	
L S L	H P L	H A A V	Q P V	A P L	
R S H C	T P S	M L L	Y N Q	. L P L	
ACCAAGAGTT	TCTATGAAGA	ACGCGGCTTC	CTGGAAATAT	TGATGCCCCA	150
Q E F	L . R	T R L P	G N I	D A P	
T K S F	Y E E	R G F	L E I L	M P H	
P R V	S M K N	A A S	W K Y	. C P I	
TCATATAGGA	GTTTATCTAA	GGGAACTCC	ACCTTCACTG	CCCACACCCA	200
S Y R S	L S K	G N S	T F T A	H T H	
H I G	V Y L R	E T P	P S L	P T P I	
I . E	F I .	G K L H	L H C	P H P	
TATGCCCCGC	AACCTGCTATA	ACTCTGCCAC	TCTTTGCATG	CATGCAAATA	250
M P R	N C Y N	S A T	L C M	H A N T	
C P A	T A I	T L P L	F A C	M Q I	
Y A P Q	L L .	L C H	S L H A	C K Y	
CTCATTATTG	GACAGGGAAA	ATGATTAAATC	CTAGTTGTCC	TGGAGGACTT	300
H Y W	T G K	M I N P	S C P	G G L	
L I I G	Q G K	. L I	L V V L	E D L	
S L L	D R E N	D . S	. L S	W R T W	
GGAGCCACTG	TCTGTGAC	TTACTTCACC	CATACCAGTA	TGTCTGATGG	350
G A T V	C W T	Y F T	H T S M	S D G	
E P L	S V G L	T S P	I P V	C L M G	
S H C	L L D	L L H P	Y Q Y	V . W	

FIG 4 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ACCTCACCTG	TGTAAAATTT	AGCAATACTA	TAGACACAAC	CAGCTCCCAA	750
L T C V K F	S N T I	D T T	S S Q		
T S P V .	N L A I L	. T Q P	A P N		
P H L C K I .	Q Y Y	R H N	Q L P M		
TGCATCAGGT	GGGTAAACACC	TCCCACACGA	ATAGTCTGCC	TACCCCTCAGG	800
C I R W V T P	P T R	I V C L	P S G		
A S G G .	H L P H E	. S A	Y P Q E		
H Q V G N T	S H T N	S L P	T L R		
AATATTTTTT	GTCGTGGGTA	CCTCAGCCTA	TCATTGTTTG	AATGGCTCTT	850
I F F V C G T	S A Y	H C L	N G S S		
Y F L S V V	P Q P I	I V .	M A L		
N I F C L W Y	L S L	S L F E	W L F		
CAGAATCTAT	GTCCTTCCTC	TCATTCTTAG	TGCCCCCTAT	GACCATCTAC	900
E S M C F L	S F L V	P P M	T I Y		
Q N L C A S S	H S .	C P L .	P S T		
R I Y V L P L	I L S	A P Y	D H L H		
ACTGAACAAG	ATTTATACAA	TCATGTGCTA	CCTAAGCCCC	ACAACAAAAG	950
T E Q D L Y N	H V V	P K P H	N K R		
L N K I Y T I	M S Y	L S P	T T K E		
. T R F I Q	S C R T	. A P	Q Q K		
AGTACCCATT	CTTCCTTTTG	TTATCAGAGC	AGGAGTGCTA	GGCAGACTAG	1000
V P I L P F V	I R A	G V L	G R L G		
Y P F F L L	L S E Q	E C .	A D .		
S T H S S F C	Y Q S	R S A R	Q T R		
GTACTGGCAT	TGGCAGTATC	ACAACCTCTA	CTCAGTTCTA	CTACAAACTA	1050
T G I G S I	T T S T	Q F Y	Y K L		
V L A L A V S	Q P L	L S S T	T N Y		
Y W H W Q Y H	N L Y	S V L	L Q T I		

FIG 4

(continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCTCAAGAAA	TAAATGGTGA	CATGGAACAG	GTCACTGACT	CCCTGGTCAC	1100
S Q E I	N G D	M E Q	V T D S	L V T	
L K K	. M V T	W N R	S L T	P W S P	
S R N	K W .	H G T G	H . L	P G H	
CTTGCAAGAT	CAACTTAACT	CCCTAGCAGC	AGTAGTCCTT	CAAAATCGAA	1150
L Q D	Q L N S	L A A	V V L	Q N R R	
C K I	N L T	P . Q Q	. S F	K I E	
L A R S	T . L	P S S	S S P S	K S K	
GAGCTTTAGA	CTTGCTAACC	GCCAAAAGAG	GGGGAACCTG	TTTATTTTIA	1200
A L D	L L T	A K R G	G T C	L F L	
E L .	T C .	P P K E	G E P V	Y F .	
S F R	L A N R	Q K R	G N L	F I F R	
GGAGAAGAAC	GCTGTATTA	TGTTAATCAA	TCCAGAATTG	TCACTGAGAA	1250
G E E R	C Y Y	V N Q	S R I V	T E K	
E K N	A V I M	L I N	P E L	S L R K	
R R T	L L L	C . S I	Q N C	H . E	
AGTTAAAGAA	ATTGAGATC	GAATACAATG	TAGAGCAGAG	GAGCTTCAAA	1300
V K E	I R D R	I Q C	R A E	E L Q N	
L K K	F E I	E Y N V	E Q R	S F K	
S . R N	S R S	N T M	. S R G	A S K	
ACACCGAACG	CTGGGGCCTC	CTCAGCCAAT	GGATGCCCTG	GGTCTCCCC	1350
T E R	W G L	L S Q W	M P W	V L P	
T P N A	G A S	S A N	G C P G	F S P	
H R T	L G P P	Q P M	D A L	G S P L	
TTCTTAGGAC	CTCTAGCAGC	TCTAATATTG	TTACTCCTCT	TTGGACCOCTG	1400
F L G P	L A A	L I L	L L L F	G P C	
S . D	L . Q L	. Y C	Y S S	L D P V	
L R T	S S S	S N I V	T P L	W T L	

FIG 4

(continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TATCTTTAAC	CTCCTTGTTA	AGTTTGCTC	TTCCAGAATT	GAAGCTGTAA
I F N	L L V K	F V S	S R I	E A V K
S L T	S L L	S L S L	P E L	K L .
Y L .	P P C .	V C L	F Q N .	S C K
AGCTACAGAT	GGTCTTACAA	ATGGAACCCC	A	1481
L Q M	V L Q	M E P		
S Y R W	S Y K	W N P		
A T D	G L T N	G T P		

FIG 5

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCAAAATCGA	AGAGCTTTAG	ACTTGCTAAC	CGCCAAAAGA	GGGGGAACCT	50
S K S K	S F R	L A N	R Q K R	G N L	
Q N R	R A L D	L L T	A K R	G G T C	
K I E	E L .	T C .	P P K E	G E P	
GTTTATTTTT	AGGGGAAGAA	TGCTGTTAGT	ATGTTAATCA	ATCTGGAATC	100
F I F	R G R M	L L V C	. S	I W N H	
L F L	G E E	C C .	Y V N Q	S G I	
V Y F .	G K N	A V S	M L I N	L E S	
ATTACTGAGA	AAGTTAAAGA	AATTGAGAT	CGAATATAAT	GTAGAGCAGA	150
Y . E	S . R	N L R S	N I M .	S R	
I T E K	V K E	I . D	R I . C	R A E	
L L R	K L K K	F E I	E Y N	V E Q R	
GGACCTTCAA	AACACTGCAC	CCTGGGGCCT	CCTCAGCCAA	TGGATGCCCT	200
G P S K	H C T	L G P	P Q P M	D A L	
D L Q	N T A P	W G L	L S Q	W M P W	
T F K	T L H	P G A S	S A N	G C P	
GGACTCTCCC	CTTCTTAGGA	CCTCTAGCAG	CTATAATATT	TTTACTCCTC	250
D S P	L L R T	S S S	Y N I	F T P L	
T L P	F L G	P L A A	I I F	L L L	
G L S P	S . D	L . Q	L . Y F	Y S S	
TTTGGACCCCT	GTATCTTCAA	CTTCCTTGTT	AAGTTTGTCT	CTTCCAGAAT	300
W T L	Y L Q	L P C .	V C L	F Q N	
F G P C	I F N	F L V	K F V S	S R I	
L D P	V S S T	S L L	S L S	L P E L	
TGAAGCTGTA	AAGCTACAAA	TAGTTCTTCA	AATGGAACCC	CAGATGCAGT	350
. S C K	A T N	S S S	N G T P	D A V	
E A V	K L Q I	V L Q	M E P	Q M Q S	
K L .	S Y K .	F F K	W N P	R C S	

FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCATGACTAA	AATCTACCGT	GGACCCCTGG	ACCGGCGCTGC	TAGACTATGC	400
H D .	N L P W	T P G	P A C .	T M L	
M T K	I Y R	G P L D	R P A	R L C	
P . L K	S T V	D P W	T G L L	D Y A	
TCTGATGTTA	ATGACATTGA	AGTCACCCCT	CCCGAGGAAA	TCTCAACTGC	450
. C . .	H .	S H P S	R G N	L N C	
S D V N	D I E	V T P	P E E I	S T A	
L M L	M T L K	S P L	P R K	S Q L H	
ACAACCCCTA	CTACACTCCA	ATTCAGTAGG	AAGCAGTTAG	AGCAGTTGTC	500
T T P T	T L Q	F S R	K Q L E	Q L S	
Q P L	L H S N	S V G	S S .	S S C Q	
N P Y	Y T P	I Q .	E A V R	A V V	
AGCCAACCTC	CCCAACAGTA	CTTGGGTTTT	CCTGTTGAGA	GGGTGGACTG	550
A N L	P N S T	W V F	L L R	G W T E	
P T S	P T V	L G F S	C . E	G G L	
S Q P P	Q Q Y	L G F	P V E R	V D .	
AGAGACAGGA	CTAGCTGGAT	TTCCTAGGCT	GACTAAGAAT	CCCAAGCCT	600
R Q D .	L D F L G .	L R I	P K P		
R D R T	S W I	S . A	D . E S	X S L	
E T G	L A G F	P R L	T K N	P X A X	
ANCTGGGAAG	GTGACCGCAT	CCATCTTTAA	ACATGGGGCT	TGCAACTTAG	650
X W E G	D R I	H L .	T W G L	Q L S	
X G K	V T A S	I F K	H G A	C N L A	
L G R .	P H	P S L N	M G L	A T .	
CTCACACCCG	ACCAATCAGA	GAGCTCACTA	AAATGCTAAT	CAGGCAAAAA	700
S H P	T N Q R	A H .	N A N	Q A K T	
H T R	P I R	E L T K	M L I	R Q K	
L T P D	Q S E	S S L	K C .	S G K N	

FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAGGAGGTAA	AGCAATAGCC	AATCATCTAT	TGCCTGAGAG	CACAGCGGGA	750
G G K	A I A	N H L L	P E S	T A G	
Q E V K	Q . P	I I Y	C L R A	Q R E	
R R .	S N S Q	S S I	A . E	H S G K	
AGGACAAGGA	TTGGGATATA	AACTCAGGCA	TTCAAGCCAG	CAACAGCAAC	800
R T R I	G I .	T Q A	F K P A	T A T	
G Q G	L G Y K	L R H	S S Q	Q Q Q P	
D K D	W D I	N S G I	Q A S	N S N	
CCCCTTTGGG	TCCCCGCCCA	TTGTATGGGA	GCTCTGTTTT	CACTCTATTT	850
P F G	S P P I	V W E	L C F	H S I S	
P L G	P L P	L Y G S	S V F	T L F	
P L W V	P S H	C M G	A L F S	L Y F	
CACTCTATTA	AATCATGCAA	CTGCACTCTT	CTGGTCCGIG	TTTTTTATGG	900
L Y .	I M Q	L H S S	G P C	F L W	
H S I K	S C N	C T L	L V R V	F Y G	
T L L	N H A T	A L F	W S V	F F M A	
CTCAAGCTGA	GCTTTTGTTT	GCCATCCACC	ACTGCTGTTT	GCCACCGTCA	950
L K L S	F C S	P S T	T A V C	H R H	
S S .	A F V R	H P P	L L F	A T V T	
Q A E	L L F	A I H H	C C L	P P S	
CAGACCCGCT	GCTGACTTCC	ATCCCTTTGG	ATCCAGCAGA	GTGTCCACTG	1000
R P A	A D F H	P F G	S S R	V S T V	
D P L	L T S	I P L D	P A E	C P L	
Q T R C	. L P	S L W	I Q Q S	V H C	
TGCTCCTGAT	CCAGCGAGGT	ACCCATTGCC	ACTCCCGATC	AGGCTAAAGG	1050
L L I	Q R G	T H C H	S R S	G . R	
C S .	S S E V	P I A	T P D Q	A K G	
A P D	P A R Y	P L P	L P I	R L K A	

FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTTGCCATTG	TTCTGTCATG	GCTAAGTGCC	TGGGTTTGTC	CTAATAGAAC	1100
L A I V	P A W	L S A	W V C P	N R T	
L P L	F L H G	. V P	G F V	L I E L	
C H C	S C M	A K C L	G L S	. . N	
TGAACACTGG	TCACTGGGTT	CCATGGTICT	CTTCCATGAC	CCACGGCTTC	1150
E H W	S L G S	M V L	F H D	P R L L	
N T G	H W V	P W F S	S M T	H G F	
. T L V	T G F	H G S	L P .	P T A S	
TAATAGAGCT	ATAACACTCA	CCGCATGGCC	CAAGATTCCA	TTCTTTGGTA	1200
I E L	. H S	P H G P	R F H	S L V	
. . S Y	N T H	R M A	Q D S I	P W Y	
N R A	I T L T	A W P	K I P	F L G I	
TCTGTGAGGC	CAAGAACCCC	AGGTCAGAGA	ANGTGAGGCT	TGCCACCATT	1250
S V R P	R T P	G Q R X	. G L	P P F	
L . G	Q E P Q	V R E X	E A	C H H L	
C E A	K N P	R S E X	V R L	A T I	
TGGGAAGTGG	CCCACTGCCA	TTTTGGTAGC	GGCCACCAC	CATCTTGGGA	1300
G K W	P T A I	L V A	A H H	H L G S	
G S G	P L P	F W .	R P T T	I L G	
W E V A	H C H	F G S	G P P P	S W E	
GCTGTGGGAG	CAAGGATCCC	CCAGTAACA			1329
C G S	K D P	P V T			
A V G A	R I P	Q .			
L W E	Q G S P	S N			

FIG 6

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCTAGAACGT	ATTCTGGAGA	ATTGGGACCA	ATGTGACACT	CAGACGCTAA	50
P R T Y	S G E	L G P	M . H S	D A K	
L E R	I L E N	W D Q	C D T	Q T L R	
. N V	F W R	I G T N	V T L	R R .	
GAAAGAAACG	ATTTATATTTC	TTCTGCAGTA	CCGCCTGGCC	ACAATATCCT	100
K E T	I Y I L	L Q Y	R L A	T I S S	
K K R	F I F	F C S T	A W P	Q Y P	
E R N D	L Y S	S A V	P P G H	N I L	
CTTCAAGGGA	GAGAAACCTG	GCTTCCTGAG	GGAAGTATAA	ATTATAACAT	150
S R E	R N L A S	. G K Y K	L . H		
L Q G R	E T W	L P E	G S I N	Y N I	
F K G	E K P G	F L R	E V .	I I T S	
CATCTTACAG	CTAGACCTCT	TCTGTAGAAA	CGAGGGCAAA	TGGAGTCAAG	200
H L T A	R P L L .	K G G Q M	E . S		
I L Q	L D L F	C R K	E G K	W S E V	
S Y S	. T S	S V E R	R A N	G V K	
TGCCATATGT	GCAAACCTTC	TTTTTCATTAA	GAGACAACCTC	ACAATTATGT	250
A I C	A N F L	F I K	R Q L	T I M .	
P Y V	Q T F	F S L R	D N S	Q L C	
C H M C	K L S	F H .	E T T H	N Y V	
AAAAAGTGIG	GTITATGCCC	TACAGGAAGC	CCTCAGAGTC	CACCTCCCTA	300
K V W	F M P	Y R K P	S E S	T S L	
K K C G	L C P	T G S	P Q S P	P P Y	
K S V	V Y A L	Q E A	L R V	H L P T	
CCCCAGCGTC	CCCTCCCCGA	CTCCTTCCTC	AACTAATAAG	GACCCCCCTT	350
P Q R P	L P D	S F L N .	. G P P F		
P S V	P S P T	P S S	T N K	D P P L	
P A S	P P R	L L P Q	L I R	T P L	
TAACCCAAAC	GGTCCAAAAG	GAGATAGACA	AAGGGGTAAA	CAATGAACCA	400
N P N	G P K G	D R Q	R G K Q .	T K	
T Q T	V Q K	E I D K	G V N	N E P	
. P K R	S K R R .	T K G .	T M N Q		
AAGAGTGCCA	ATATTCCCCG	ATTATGCCCC	CTCCAAGCAG	TCAGAGGAGG	450
E C Q	Y S P	I M P P	P S S	E R R	
K S A N	I P R	L C P	L Q A V	R G G	
R V P	I F P D	Y A P	S K Q .	E E E	
AGAATTCCGC	CCAGCCAGAG	TGCCTGTACC	TTTTTCTCTC	TCAGACTTAA	500
R I R P	S Q S	A C T	F F S L	R L K	
E F G	P A R V	P V P	F S L	S D L K	
N S A	Q P E	C L Y L	F L S	Q T .	

FIG 6 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGCAAATTAA	AATAGACCTA	GGTAAATTCT	CAGATAACCC	TGACGGCTAT	550
A N .	N R P R .	I L R .	P .	R L Y	
Q I K	I D L	G K F S	D N P	D G Y	
S K L K .	T .	V N S	Q I T L	T A I	
ATTGATGTTT	TACAAGGGTT	AGGACAATCC	TTTGATCTGA	CATGCAGAGA	600
. C F	T R V	R T I L	. S D	M E R	
I D V L	Q G L	G Q S	F D L T	W R D	
L M F	Y K G .	D N P	L I .	H G E I	
TATAATGTTA	CTACTAAATC	AGACACTAAC	CCCAAATGAG	AGAAGTGCCG	650
Y N V T	T K S	D T N	P K .	E K C R	
I M L	L L N Q	T L T	P N E	R S A A	
. C Y	Y .	I R H .	P Q M R	E V P	
CTGTAAGTGC	AGCCCCGAGAG	TTTGGCGATC	TTTGGTATCT	CAGTCAGGCC	700
C N C	S P R V	W R S	L V S	Q S G Q	
V T A	A R E	F G D L	W Y L	S Q A	
L .	L Q	P E S	L A I	F G I S	V R P
AACAATAGGA	TGACAACAGA	GGAAAGAACA	ACTCCACAG	GCCAGCAGGC	750
Q .	D D N R	G K N N	S H R	P A G	
N N R M	T T E	E R T	T P T G	Q Q A	
T I G	. Q Q R	K E Q	L P Q	A S R Q	
AGTTOCCAGT	GTAGACCCTC	ATTGGGACAC	AGAATCAGAA	CATGCAGATT	800
S S Q C	R P S	L G H	R I R T	W R L	
V P S	V D P H	W D T	E S E	H G D W	
F P V	. T L	I G T Q	N Q N	M E I	
GGTGCCACAA	ACATTGCTA	ACTTGGTGC	TAGAAGGACT	GAGGAAACT	850
V P Q	T F A N	L R A	R R T	E E N .	
C H K	H L L	T C V L	E G L	R K T	
G A T N	I C .	L A C	. K D .	G K L	
AGGAAGAAGC	CTATGAATTA	CTCAATGATG	TCCACTATAA	CACAGGGAAA	900
E E A	Y E L	L N D V	H Y N	T G K	
R K K P	M N Y	S M M	S T I T	Q G K	
G R S	L .	I T Q .	C P L .	H R E R	
GGAAGAAAAT	CTTACTGCTT	TTCTGGACAG	ACTAAGGGAG	GCATTGAGGA	950
G R K S	Y C F	S G Q	T K G G	I E E	
E E N	L T A F	L D R	L R E	A L R K	
K K I	L L L	F W T D	. G R H .	G	
AGCATACCTC	CCTGTACCT	GACTCTATTG	AAGGCCAACT	AATCTTAAAG	1000
A Y L	P V T .	L Y .	R P T	N L K G	
H T S	L S P	D S I E	G Q L	I L K	
S I P P	C H L	T L L	K A N .	S . R	

FIG 6 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GATAAGTTTA	TCACTCAGTC	ACCTGCAGAC	ATTAGAAAAA	ACTTCAAAAG	1050
. V Y H S V	S C R H	. K K L Q K			
D K F I T Q S	A A D I R K N	F K S			
I S L S L S Q	L Q T L E K	T S K V			
TCTGCCTTAG	GCCCCGAGCA	GAACCTAGAA	ACCCTATTTA	ACTTGGCATC	1100
S A L G P E Q	N L E T L F N	L A S			
L P . A R S R	T . K P Y L	T W H P			
C L R P G A	E L R N P I	. L G I			
CTCAGTTTTT	TATAATAGAG	ATCAGGAGGA	GCAGGCGAAA	CGGACAAAC	1150
S V F Y N R D	Q E E Q A K	R D K R			
Q F F I I E	I R R S R R N	G T N			
L S F L . . R	S G G A G E T	G Q T			
GGGATAAAAA	AAAAAGGGG	GGTCCACTAC	TTTAGTCATG	GCCCTCAGGC	1200
D K K K R G	G P L L . S W	P S G			
G I K K K G G	V H Y F S H G	P Q A			
G . K K K G G	S T T L V M	A L R Q			
AAGCAGACTT	TGAGGCTCT	GCAAAAGGGA	AAAGCTGGGC	AAATCAAATG	1250
K Q T L E A L	Q K G K A G Q	I K C			
S R L W R L C	K R E K L G	K S N A			
A D F G G S	A K G K S W A	N Q M			
OCTAATAGGG	CTGGCTTCCA	GTGCGGTCTA	CAAGGACACT	TTAAAAAAGA	1300
L I G L A S S	A V Y K D T	L K K I			
. . G W L P	V R S T R T L	. K R			
P N R A G F Q	C G L Q G H F	K K D			
TTATCCAAGT	AGAAATAAGC	CGCCCCCTTG	TCCATGCCCC	TTAGTCAAG	1350
I Q V E I S	R P L V H A P	Y V K			
L S K . K . A	A P L S M P L	T S R			
Y P S R N K P	P P C P C P	L R Q G			
GGAATCACTG	GAAGGCCAC	TGCCCCAGGG	GATGAAGATA	CTCTGAGTCA	1400
G I T G R P T	A P G D E D T	L S Q			
E S L E G P L	P Q G M K I	L . V R			
N H W K A H	C P R G . R Y	S E S			
GAAGCCATTA	ACCAGATGAT	CCAGCAGCAG	GAATGAGGGT	GCCCCGGGGC	1450
K P L T R . S	S S R T E G	A R G E			
S H . P D D	P A A G L R V	P G A			
E A I N Q M I	Q Q Q D . G C	P G R			
AGGCCAGCC	CATGCCATCA	CCCTCACAGA	GCCCCGGGTA	TGTTTGACCA	1500
R Q P M P S	P S Q S P G Y	V . P			
S A S P C H H	P H R A P G M	F D H			
A P A H A I T	L T E P R V	C L T I			

FIG 6

(continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TTCAGAGCCA A				
1511				
L	R	A		
	E	P		
E	S	Q		

FIG 7

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGGCAGCA	GCCATCATCA	TCATCATCAC	AGCAGCGGCC	TGGTGCCGCG	50
M G S S	H H H	H H H	S S G L	V P R	
CGGCAGCCAT	ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTCGGATCC	100
G S H	M A S M	T G G	Q Q M	G R I L	
TAGAACGTAT	TCTGGAGAAT	TGGGACCAAT	GTGACACTCA	GACGCTAAGA	150
E R I	L E N	W D Q C	D T Q	T L R	
AAGAAACGAT	TTATATTCTT	CTGCAGTACC	GCCTGGCCAC	AATATCCTCT	200
K K R F	I F F	C S T	A W P Q	Y P L	
TCAAGGGAGA	GAAACCTGGC	TTCCTGAGGG	AAGTATAAAT	TATAACATCA	250
Q G R	E T W L	P E G	S I N	Y N I I	
TCTTACAGCT	AGACCTCTTC	TGTAGAAAGG	AGGGCAAATG	GAGTGAAGTG	300
L Q L	D L F	C R K E	G K W	S E V	
CCATATGTGC	AAACTTTCTT	TTCATTAAGA	GACAACTCAC	AATTATGTAA	350
P Y V Q	T F F	S L R	D N S Q	L C K	
AAAGTGTGGT	TTATGCCCTA	CAGGAAGCCC	TCAGAGTCCA	OCTCCCTACC	400
K C G	L C P T	G S P	Q S P	P P Y P	
CCAGCGTCCC	CTCCCCGACT	OCTTCCTCAA	CTAATAAGGA	CCCCCCTTTA	450
S V P	S P T	P S S T	N K D	P P L	
ACCCAAACGG	TCCAAAAGGA	GATAGACAAA	GGGTAAACA	ATGAACCAAA	500
T Q T V	Q K E	I D K	G V N N	E P K	
GAGTGCCAAT	ATTCCCGGAT	TATGCCCCCT	CCAAGCAGTG	AGAGGAGGAG	550
S A N	I P R L	C P L	Q A V	R G G E	
AATTGGGCCC	AGCCAGAGTG	CCTGTACCTT	TTCTCTCTCT	AGACTTAAAG	600
F G P	A R V	P V P F	S L S	D L K	
CAAATTAAAA	TAGACCTAGG	TAAATTCTCA	GATAACCCCTG	ACGGCTATAT	650
Q I K I	D L G	K F S	D N P D	G Y I	
TGATGTTTTA	CAAGGGTTAG	GACAATCCTT	TGATCTGACA	TGGAGAGATA	700
D V L	Q G L G	Q S F	D L T	W R D I	
TAATGTTACT	ACTAAATCAG	AACTAAACCC	CAAATGAGAG	AAGTGCCGCT	750
M L L	L N Q	T L T P	N E R	S A A	
GTAAGTGCAG	CCCGAGAGTT	TGGCGATCTT	TGGTATCTCA	GTCAGGCCAA	800
V T A A	R E F	G D L	W Y L S	Q A N	

FIG 7 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAATAGGATG	ACAACAGAGG	AAAGAACAAC	TCCACAGGC	CAGCAGGCAG	850
N R M	T T E E	R T T	P T G	Q Q A V	
TTCCAGTGT	AGACCCTCAT	TGGGACACAG	AATCAGAACA	TGGAGATTGG	900
P S V	D P H	W D T E	S E H	G D W	
TGCCACAAAC	ATTTGCTAAC	TTGCGTGCTA	GAAGGACTGA	GGAAAACTAG	950
C H K H	L L T	C V L	E G L R	K T R	
GAAGAAGCCT	ATGAATTACT	CAATGATGTC	CACTATAACA	CAGGGAAAGG	1000
K K P	M N Y S	M M S	T I T	Q G K E	
AAGAAAATCT	TACTGCTTTT	CTGGACAGAC	TAAGGGAGGC	ATTGAGGAAG	1050
E N L	T A F	L D R L	R E A	L R K	
CATACCTCCC	TGTCACCTGA	CTCTATTGAA	GGCCAACTAA	TCTTAAAGGA	1100
H T S L	S P D	S I E	G Q L I	L K D	
TAAGTTTATC	ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAC	TTCAAAAGTC	1150
K F I	T Q S A	A D I	R K N	F K S L	
TGCCTAAGCT	TGCGGCCGCA	CTCGAGCACC	ACCACCACCA	CCACTGAGAT	1200
P K L	A A A	L E H H	H H H	H . D	
CCGGCTGCTA	ACAAAGCCCG	AAAGGAAGCT	GAGTTGGCTN	GTGGCNA	1247
P A A N	K A R	K E A	E L A X	G	

FIG 8

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTCGGATCC	TAGAACGTAT	50
M A S M	T G G	Q Q M	G R I L	E R I	
TCTGGAGAAT	TGGGACCAAT	GTGACACTCA	GACGCTAAGA	AAGAAACGAT	100
L E N	W D Q C	D T Q	T L R	K K R F	
TTATATTCTT	CTGCAGTACC	GCCTGGCCAC	AATATCCTCT	TCAAGGGAGA	150
I F F	C S T	A W P Q	Y P L	Q G R	
GAAACCTGGC	TTCTTGAGGG	AAGTATAAAT	TATAACATCA	TCTTACAGCT	200
E T W L	P E G	S I N	Y N I I	L Q L	
AGACCTCTTC	TGTAGAAAGG	AGGGCAAATG	GAGTGAAGTG	CCATATGTGC	250
D L F	C R K E	G K W	S E V	P Y V Q	
AAACTTTCTT	TTCATTAGA	GACAACTCAC	AATTATGTAA	AAAGTGIGGT	300
T F F	S L R	D N S Q	L C K	K C G	
TTATGCCCTA	CAGGAAGCCC	TCAGAGTCCA	CCTCCCTACC	CCAGCGTCCC	350
L C P T	G S P	Q S P	P P Y P	S V P	
CTCCCCGACT	CCTTCCTCAA	CTAATAAGGA	CCCCCCTTTA	ACCCAAACGG	400
S P T	P S S T	N K D	P P L	T Q T V	
TCCAAAAGGA	GATAGACAAA	GGGGTAAACA	ATGAACCAAA	GAGTGCCAAT	450
Q K E	I D K	G V N N	E P K	S A N	
ATTCCCCGAT	TATGCCCCCT	CCAAGCAGTG	AGAGGAGGAG	AATTGGCCCC	500
I P R L	C P L	Q A V	R G G E	F G P	
AGCCAGAGTG	CCTGTACCTT	TTTCTCTCTC	AGACTTAAAG	CAAATTAAAA	550
A R V	P V P F	S L S	D L K	Q I K I	
TAGACCTAGG	TAAATTCTCA	GATAACCCCTG	ACGGCTATAT	TGATGTTTTA	600
D L G	K F S	D N P D	G Y I	D V L	
CAAGGGTTAG	GACAATCCTT	TGATCTGACA	TGGAGAGATA	TAATGTTACT	650
Q G L G	Q S F	D L T	W R D I	M L L	
ACTAAATCAG	ACACTAACCC	CAAATGAGAG	AAGTGCCGCT	GTAAGTGCAG	700
L N Q	T L T P	N E R	S A A	V T A A	
CCCGAGAGTT	TGGCGATCTT	TGGTATCTCA	GTCAGGCCAA	CAATAGGATG	750
R E F	G D L	W Y L S	Q A N	N R M	
ACAACAGAGG	AAAGAACAAC	TCCCACAGGC	CAGCAGGCAG	TTCCCAGTGT	800
T T E E	R T T	P T G	Q Q A V	P S V	

FIG 8 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGACCCATCAT	TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCCACAAAC	850
D P H	W D T E	S E H	G D W	C H K H	
ATTTGCTAAC	TTGCGTGCTA	GAAGGACTGA	GGAAACTAG	GAAGAAGCCT	900
L L T	C V L	E G L R	K T R	K K P	
ATGAATTACT	CAATGATGTC	CACTATAACA	CAGGGAAAGG	AAGAAAATCT	950
M N Y S	M M S	T I T	Q G K E	E N L	
TACTGCTTTT	CTGGACAGAC	TAAGGGAGGC	ATTGAGGAAG	CATACCTCCC	1000
T A F	L D R L	R E A	L R K	H T S L	
TGTCACCTGA	CTCTATTGAA	GGCCAACTAA	TCTTAAAGGA	TAAGTTTATC	1050
S P D	S I E	G Q L I	L K D	K F I	
ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAC	TTCAAAAGTC	TGCCTAAGCT	1100
T Q S A	A D I	R K N	F K S L	P K L	
TGCGGCCGCA	CTCGAGCAAC	ACCACCACCA	CCACTGAGAT	CCGGCTGCTA	1150
A A A	L E H H	H H H	H . D	P A A N	
ACAAAGCCCCG	AAAGGAAGCT	GAGTTGGCTG	GTGGCA		1186
K A R	K E A	E L A G	G		

FIG 9

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTGCC	TCTCCCAATT	50
C P L C	S . S	S T G	A H C L	S Q L	
V R C	A P D P	A Q A	P I A	S P N W	
S A V	L L I	Q H R R	P L P	L P I	
GGGCTAAAGG	CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTTCATC	100
G . R	L A I V	P A Q	L S A	W V H P	
A K G	L P L	F L H S	. V P	G F I	
G L K A	C H C	S C T	A K C L	G S S	
CTAATCGAGC	TGAACACTAG	TCACTGGGTT	CCACGGTTC	CTTCATGAC	150
N R A	E H .	S L G S	T V L	F H D	
L I E L	N T S	H W V	P R F S	S M T	
. S S	. T L V	T G F	H G S	L P . P	
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L	I E L	. H S	L H G P	R F H	
H G F	. . S Y	N T H	C M V	Q D S I	
M A S	N R A	I T L T	A W S	K I P	
TTCCTTGGAA	TCCGIGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E	S V R P	R T P	G Q R	T Q G L	
P W N	P . D	Q E P Q	V R E	H K A	
F L G I	R E T	K N P	R S E N	T R L	
TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTTGGAAGC	AGCCCGCCAC	300
P P C	W K Q	P T T I	L E A	A R H	
C H H V	G S S	P P P	F W K Q	P A T	
A T M	L E A A	H H H	F G S	S P P L	
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	AGGTAACAAT	TTGGTGACCA	350
Y L G S	S G S	K D P	R . Q F	G D H	
I L G	A L G A	R T P	G N N	L V T T	
S W E	L W E	Q G P Q	V T I	W . P	
CGAAGGGACC	TGAATCCGCA	ACCATGAAGG	GATCTCCAAA	GCAATTGGAA	400
E G T	. I R N	H E G	I S K	A I G N	
K G P	E S A	T M K G	S P K	Q L E	
R R D L	N P Q	P . R	D L Q S	N W K	

FIG 9

(continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGTTCCCTCC	CAAGGCAAAA	ATGCCCCCTAA	GATGTATTCT	GGAGAATTGG	450
V P P	K A K	M P L R	C I L	E N W	
M F L P	R Q K	C P .	D V F W	R I G	
C S S	Q G K N	A P K	M Y S	G E L G	
GACCAATTTG	ACCCTCAGAC	AGTAAGAAAA	AAATGACTTA	TATTCTTCTG	500
D Q F D	P Q T	V R K	K . L I	F F C	
T N L	T L R Q	. E K	N D L	Y S S A	
P I .	P S D	S K K K	M T Y	I L L	
CAGTACCGCC	CTGGCCACGA	TATCCTCTTC	AAGGGGGAGA	AACCTGGCCT	550
S T A	L A T I	S S S	R G R	N L A S	
V P P	W P R	Y P L Q	G G E	T W P	
Q Y R P	G H D	I L F	K G E K	P G L	
CCTGAGGGAA	GTATAAATTA	TAACACCATC	TTACAGCTAG	ACCTGTTTTG	600
. G K	Y K L	. H H L	T A R	P V L	
P E G S	I N Y	N T I	L Q L D	L F C	
L R E V	. I I	T P S	Y S .	T C F V	
TAGAAAAGGA	GGCAAATGGA	GTGAAGTGCC	ATATTTACAA	ACTTTCCTTT	650
. K R R	Q M E	. S A	I F T N	F L F	
R K G	G K W S	E V P	Y L Q	T F F S	
E K E	A N G	V K C H	I Y K	L S F	
CATTAAAAGA	CAACTCGCAA	TTATGTTAAC	AGTGTGATTT	GIGTTCCTAC	700
I K R	Q L A I	M L T	V . F	V F L H	
L K D	N S Q	L C .	Q C D L	C S Y	
H . K T	T R N	Y V N	S V I C	V P T	
ACGGAAGCCC	TCAGATTCTA	CTCCCCACCC	CCGGCATCTC	CCCTGAATCC	750
G S P	Q I L	L P T P	G I S	P E S	
T E A L	R F Y	S P P	P A S P	L N P	
R K P	S D S T	P H P	R H L	P . I P	
CTCCCCAACT	TATT				764
L P N L					
S P T Y					
P Q L I					

FIG 10

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCCTGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTTGCC	TCTCCCAATT	50
C P L C	S . S	S T G	A H C L	S Q L	
V R C	A P D P	A Q A	P I A	S P N W	
S A V	L L I	Q H R R	P L P	L P I	
GGGCTAAAGG	CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTCATC	100
G . R	L A I V	P A Q	L S A	W V H P	
A K G	L P L	F L H S	. V P	G F I	
G L K A	C H C	S C T	A K C L	G S S	
CTAATCGAGC	TGAACACTAG	TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	150
N R A	E H .	S L G S	T V L	F H D	
L I E L	N T S	H W V	P R F S	S M T	
. S S	. T L V	T G F	H G S	L P . P	
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L	I E L .	H S	L H G P	R F H	
H G F	. . S Y	N T H	C M V	Q D S I	
M A S	N R A	I T L T	A W S	K I P	
TTCCTTGGAA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E	S V R P	R T P	G Q R	T Q G L	
P W N	P . D	Q E P Q	V R E	H K A	
F L G I	R E T	K N P	R S E N	T R L	
TGCCACCATG	TTGGAAGCAG	CCCACCAcca	TTTTGGAAGC	GGCCCCGCCAC	300
P P C	W K Q	P T T I	L E A	A R H	
C H H V	G S S	P P P	F W K R	P A T	
A T M	L E A A	H H H	F G S	G P P L	
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	CAGGTAACAA	TTTGGTGACC	350
Y L G S	S G S	K D P	Q V T I	W . P	
I L G	A L G A	R T P	R . Q	F G D H	
S W E	L W E	Q G P P	G N N	L V T	
ACGAAGGGAC	CTGAATCCGC	AACCATGAAG	GGATCTCCAA	AGCAATTGGA	400
R R D	L N P Q	P . R	D L Q	S N W K	
E G T	. I R	N H E G	I S K	A I G	
T K G P	E S A	T M K	G S P K	Q L E	

FIG 10

(continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AATGTTCTC	CCAAGGCAA	AATGCCCCA	AGATGTATC	TGGAGAATTG	450
C S S	Q G K	N A P K	M Y S	G E L	
N V P P	K A K	M P L	R C I L	E N W	
M F L	P R Q K	C P .	D V F	W R I G	
GGACCAATCT	GACCTCAGA	CAGTAAGAA	AAAAATGACT	TATATCTTC	500
G P I .	P S D	S K K	K N D L	Y S S	
D Q S	D P Q T	V R K	K M T	Y I L L	
T N L	T L R	Q .	E K	K . L I F F	
TGCAGTACCG	CCTGGCCACG	GATATCTCT	TCAAGGGGA	GAAACCTGGC	550
A V P	P G H G	Y P L	Q G G	E T W P	
Q Y R	L A T	D I L F	K G E	K P G	
C S T A	W P R	I S S	S R G R	N L A	
CTCTGAGGG	AAGTATAAAT	TATAACACCA	TCTTACAGCT	AGACCTGTTT	600
P E G	S I N	Y N T I	L Q L	D L F	
L L R E	V .	I I T P	S Y S .	T C F	
S .	G K Y K L	. H H	L T A	R P V L	
TGTAGAAAAG	GAGGCAAATG	GAGTGAAGTG	CCATATTTAC	AACTTTCTT	650
C R K G	G K W	S E V	P Y L Q	T F F	
V E K	E A N G	V K C	H I Y	K L S F	
. K R	R Q M	E .	S A	I F T N F L	
TTCATTAAAA	GACAACTCGC	AATTATGTAA	ACAGTGTGAT	TTGTGTCTTA	700
S L K	D N S Q	L C K	Q C D	L C P T	
H .	K T T R	N Y V N	S V I	C V L	
F I K R	Q L A	I M .	T V .	F V S Y	
CAGGAAGCCC	TCAGATCTAC	CTCCCTACCC	CGGCATCTCC	CTGACTCCTT	750
G S P	Q I Y	L P T P	A S P	. L L	
Q E A L	R S T	S L P	R H L P	D S F	
R K P	S D L P	P Y P	G I S	L T P S	
CCCCAACTAA	TAAGGACCCA	CTTCAGCCCA	AACAGTCCAA	AAGGACATAG	800
P Q L I	R T H	F S P	N S P K	G H	
P N .	. G P T	S A Q	T V Q	K D I	
P T N	K D P	L Q P K	Q S K	R T .	

FIG 11

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GGCATTGATA	GCACCCATCA	GATGGCCAAA	TCATTATTIA	CTGGACCAGG	50
G I D S	T H Q	M A K	S L F T	G P G	
A L I	A P I R	W P N	H Y L	L D Q A	
H . .	H P S	D G Q I	I I Y	W T R	
CCTTTTCAAA	ACTATCAAGC	AGATAGGGCC	CGTGAAGCAT	GCCAAAGAAA	100
L F K	T I K Q	I G P	V K H	A K E I	
F S K	L S S	R . G P	. S M	P K K	
P F Q N	Y Q A	D R A	R E A C	Q R N	
TAATCCCCCTG	CCTTATCGCC	ATGTTCTTTC	AGGAGAACAA	AGAACAGGCC	150
I P C	L I A	M F L Q	E N K	E Q A	
. S P A	L S P	C S F	R R T K	N R P	
N P L	P Y R H	V P S	G E Q	R T G H	
ATTACCCAGG	GGAAGACTGG	CAACTAGATT	TTACCCACAT	GGCCAAATGT	200
I T Q G	K T G	N . I	L P T W	P N V	
L P R	G R L A	T R F	Y P H	G Q M S	
Y P G	E D W	Q L D F	T H M	A K C	
CAGGGATTTC	AGCATCTACT	AGTCTGGGCA	GATACTTTCA	CTGGTITGGGT	250
R D F	S I Y .	S G Q	I L S	L V G W	
G I S	A S T	S L G R	Y F H	W L G	
Q G F Q	H L L	V W A	D T F T	G W V	
GGAGICTTCT	CCTTGTAAGGA	CAGAAAAGAC	CCAAGAGGTA	ATAAAGGCAC	300
S L L	L V G	Q K R P	K R .	. R H	
G V F S	L . D	R K D	P R G N	K G T	
E S S	P C R T	E K T	Q E V	I K A L	
TAATGAAATA	ATTCCAGAT	TTGGACTION	CCCAGGATTA	CAGGGTGACA	350
. . N N	S Q I	W T S	P R I T	G . Q	
N E I	I P R F	G L P	P G L	Q G D N	
M K .	F P D	L D F P	Q D Y	R V T	

FIG 11 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCCGC	TTTCAAGGCT	GCAGTAAACC	AGGGAGTATC	CCAGGTGTTA	400
W P R	F Q G C	S N P	G S I	P G V R	
G P A	F K A	A V T Q	G V S	Q V L	
M A P L	S R L	Q .	P R E Y P	R C .	
GGCATACAAT	ATCACTTACA	CTGTGCCTGG	AGGCCACAAT	CCTCCAGAAA	450
H T I	S L T	L C L E	A T I	L Q K	
G I Q Y	H L H	C A W	R P Q S	S R K	
A Y N	I T Y T	V P G	G H N	P P E K	
AGTCAAGAAA	ATGAATGAAA	CACTCAAAGA	TCTAAAAAAG	CTAACCCAAG	500
S Q E N	E .	N T Q R	S K K A	N P R	
V K K	M N E T	L K D	L K K	L T Q E	
S R K	. M K	H S K I	. K S	. P K	
AAACCCACAT	TGCATGACCT	GTTCTGTGTC	CTATAACCTT	ACTAAGAATC	550
N P H	C M T C	S V A	Y N L	T K N P	
T H I	A .	P V L L P	I T L	L R I	
K P T L	H D L	F C C	L .	P Y . E S	
CATAACTATC	CCCCAAAAAG	CAGGACTTAG	CCCATACGAG	ATGCTATATG	600
. L S	P K K	Q D L A	H T R	C Y M	
H N Y P	P K S	R T .	P I R D	A I W	
I T I	P Q K A	G L S	P Y E	M L Y G	
GATGGCCTTT	CCTAACCAAT	GACCTTGTC	TTGACTGAGA	AATGGCCAAC	650
D G L S	. P M	T L C	L T E K	W P T	
M A F	P N Q .	P C A	. L R	N G Q L	
W P F	L T N	D L V L	D .	E M A N	
TTAGTTGCAG	ACATCACCTC	CTTAGCCAAA	TATCAACAAG	TTCTTAA AAC	700
. L Q	T S P P	. P N	I N K	F L K H	
S C R	H H L	L S Q I	S T S	S . N	
L V A D	I T S	L A K	Y Q Q V	L K T	

FIG 11 (continued)

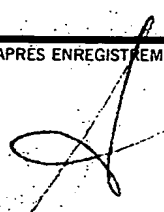
10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATCACAGGGA	ACCTGTCCCC	GAGAGGAGGG	AAAGGAACTA	TTCCACCCCTG	750
H R E	P V P	E R R E	R N Y	S T L	
I T G N	L S P	R G G	K G T I	P P W	
S Q G	T C P R	E E G	K E L	F H P G	
GTCACATG					758
V T					
. H					
D M					

NOTIFICATION DU NUMÉRO D'ENREGISTREMENT NATIONAL

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Téléphone : 01 53 04 53 04 Télécopie : 01 42 93 59 30

Confirmation d'un dépôt par télécopie ☐

A envoyer par l'INPI au demandeur ou au mandataire

DATE DE REMISE DES PIÈCES 27 JUL. 1997 N° D'ENREGISTREMENT NATIONAL 97 08816 DÉPARTEMENT DE DÉPÔT LY DATE DE DÉPÔT		1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE A QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE CABINET GERMAIN & MAUREAU B.P. 6153 69466 LYON CEDEX 06									
2 DEMANDE Nature du titre de propriété industrielle <input checked="" type="checkbox"/> brevet d'invention <input type="checkbox"/> demande divisionnaire <input type="checkbox"/> certificat d'utilité <input type="checkbox"/> transformation d'une demande de brevet européen		n° du pouvoir permanent <input type="checkbox"/> demande initiale <input type="checkbox"/> brevet d'invention <input type="checkbox"/> certificat d'utilité n° références du correspondant MD/MK/B05B 2867 <input type="checkbox"/> date 04 72 69 84 30									
Établissement du rapport de recherche <input type="checkbox"/> différé <input checked="" type="checkbox"/> immédiat Le demandeur, personne physique, requiert le paiement échelonné de la redevance <input type="checkbox"/> oui <input type="checkbox"/> non											
Titre de l'invention (200 caractères maximum) MATÉRIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES NOTAMMENT ASSOCIÉS A LA SCLEROSE EN PLAQUES ET/OU LA POLYARTHRITE RHUMATOÏDE, A DES FINS DE DIAGNOSTIC, PROPHYLACTIQUES ET THERAPEUTIQUES											
3 DEMANDEUR (S) n° SIREN <input type="text"/> code APE-NAF <input type="text"/> Nom et prénoms (souligner le nom patronymique) ou dénomination BIO MERIEUX		Forme juridique SA									
Nationalité (s) FRANCAISE Adresse (s) complète (s) CHEMIN DE L'ORME 69380 MARCY L'ETOILE		Pays FRANCE									
4 INVENTEUR (S) Les inventeurs sont les demandeurs <input type="checkbox"/> oui <input checked="" type="checkbox"/> non En cas d'insuffisance de place, poursuivre sur papier libre <input type="checkbox"/> Si la réponse est non, fournir une désignation séparée											
5 RÉDUCTION DU TAUX DES REDEVANCES <input type="checkbox"/> requise pour la 1ère fois <input type="checkbox"/> requise antérieurement au dépôt ; joindre copie de la décision d'admission											
6 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE <table border="1"> <thead> <tr> <th>pays d'origine</th> <th>numéro</th> <th>date de dépôt</th> <th>nature de la demande</th> </tr> </thead> <tbody> <tr> <td> </td> <td> </td> <td> </td> <td> </td> </tr> </tbody> </table>				pays d'origine	numéro	date de dépôt	nature de la demande				
pays d'origine	numéro	date de dépôt	nature de la demande								
7 DIVISIONS antérieures à la présente demande n° <input type="text"/> date <input type="text"/> n° <input type="text"/> date <input type="text"/>											
8 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (nom et qualité du signataire n° d'identification) Mireille DIDIER CPI		SIGNATURE DU PRÉPOSÉ À LA RÉCEPTION N. AMERIS									
SIGNATURE APRÈS ENREGISTREMENT DE LA DEMANDE À L'INPI 											

La Sclérose en Plaques (SEP) est une maladie démyélinisante du système nerveux central (SNC) dont la cause complète reste encore inconnue.

De nombreux travaux ont étayé l'hypothèse d'une
5 étiologie virale de la maladie, mais aucun des virus connus testés ne s'est avéré être l'agent causal recherché: une revue des virus recherchés depuis des années dans la SEP a été faite par E. Norrby et R.T. Johnson.

10 Récemment, un rétrovirus, différent des rétrovirus humains connus a été isolé chez des patients atteints de SEP. Les auteurs ont aussi pu montrer que ce rétrovirus pouvait être transmis in vitro, que des patients atteints de SEP produisaient des anticorps susceptibles de
15 reconnaître des protéines associées à l'infection des cellules leptoméningées par ce rétrovirus, et que l'expression de ce dernier pouvait être fortement stimulée par les gènes immédiats-précoces de certains herpesvirus.

Tous ces résultats plaident en faveur du rôle dans
20 la SEP d'au moins un rétrovirus inconnu ou d'un virus ayant une activité transcriptase inverse (RT) détectable selon la méthode publiée par H. Perron et qualifiée d'activité "RT de type LM7".

Les travaux de la Demanderesse ont permis
25 d'obtenir deux lignées continues de cellules infectées par des isolats naturels provenant de deux patients différents atteints de SEP, par un procédé de culture tel que décrit dans le document WO-A-93 20188, dont le contenu est incorporé par référence à la présente description. Ces
30 deux lignées dérivées de cellules de plexus-choroïdes humains, dénommées LM7PC et PLI-2 ont été déposées à l'E.C.A.C.C. respectivement le 22 juillet 1992 et le 8 janvier 1993, sous les numéros 92 072201 et 93 010817, conformément aux dispositions du Traité de Budapest. Par
35 ailleurs, les isolats viraux possédant une activité RT de type LM7 ont également été déposés à l'E.C.A.C.C. sous la

dénomination globale de "souches". La "souche" ou isolat hébergé par la lignée PLI-2, dénommée POL-2, a été déposée auprès de l'E.C.A.C.C. le 22 juillet 1992 sous le n° V92072202. La "souche" ou isolat hébergé par la lignée
5 LM7PC, dénommée MS7PG, a été déposée auprès de l'E.C.A.C.C. le 8 janvier 1993 sous le n° V93010816.

A partir des cultures et des isolats précités, caractérisés par des critères biologiques et morphologiques, on s'est ensuite attaché à caractériser le
10 matériel nucléaire associé aux particules virales produites dans ces cultures.

Les portions de génome déjà caractérisées ont été utilisées pour mettre au point des tests de détection moléculaire du génome viral et des tests
15 immunosérologiques, utilisant les séquences d'acides aminés codées par les séquences nucléotidiques du génome viral, pour détecter la réponse immunitaire dirigée contre des épitopes associés à l'infection et/ou l'expression virale.

Ces outils ont déjà permis de confirmer une
20 association entre la SEP et l'expression des séquences identifiées dans les brevets cités plus loin. Cependant, le système viral découvert par la Demanderesse, s'apparente à un système rétroviral complexe. En effet, les séquences retrouvées encapsidées dans les particules
25 virales extracellulaires produites par les différentes cultures de cellules de patients atteints de SEP, montrent clairement qu'il y a co-encapsidation de génomes rétroviraux apparentés, mais différents du génome rétroviral "sauvage" qui produit les particules virales
30 infectantes. Ce phénomène a été observé entre des rétrovirus répliatifs et des rétrovirus endogènes appartenant à la même famille, voire même hétérologues. La notion de rétrovirus endogène est très importante dans le contexte de notre découverte car, dans le cas de MSRV-1,
35 on a observé que des séquences rétrovirales endogènes comprenant des séquences homologues au génome MSRV-1,

existent dans l'ADN humain normal. L'existence d'éléments rétroviraux endogènes (ERV) apparentés à MSRV-1 par tout ou partie de leur génome, explique le fait que l'expression du rétrovirus MSRV-1 dans les cellules

5 humaines puisse interagir avec des séquences endogènes proches. Ces interactions sont retrouvées dans le cas de rétrovirus endogènes pathogènes et/ou infectieux (par exemple certaines souches écotropes du Murine Leukaemia virus), dans le cas de rétrovirus exogènes dont la

10 séquence nucléotidique peut être retrouvée partiellement ou en totalité, sous forme d'ERVs, dans le génome de l'animal hôte (ex. virus exogène de la tumeur mammaire de la souris transmis par le lait). Ces interactions consistent principalement en (i) une transactivation ou

15 co-activation d'ERVs par le rétrovirus répliatif, (ii) une encapsidation "illégitime" d'ARN apparentés d'ERVs, ou d'ERVs -voire d'ARN cellulaires- possédant simplement des séquences d'encapsidation compatibles, dans les particules rétrovirales produites par l'expression de la souche

20 répliatrice, parfois transmissibles et parfois avec une pathogénicité propre, et (iii) des recombinaisons plus ou moins importantes entre les génomes co-encapsidés, notamment dans les phases de transcription inverse, qui conduisent à la formation de génomes hybrides, parfois

25 transmissibles et parfois avec une pathogénicité propre.

Ainsi, (i) différentes séquences apparentées à MSRV-1 ont été retrouvées dans les particules virales purifiées; (ii) l'analyse moléculaire des différentes régions du génome rétroviral MSRV-1 doit être faite en

30 analysant systématiquement les séquences co-encapsidées, interférantes et/ou recombinaisonnelles qui sont générées par l'infection et/ou l'expression de MSRV-1, de plus, certains clones peuvent avoir des parties de séquences défectives produites par la répliation rétrovirale et les

35 erreurs de matrice et/ou de transcription de la transcriptase inverse; (iii) les familles de séquences

apparentées à une même région génomique rétrovirale sont les supports d'une détection diagnostique globale qui peut être optimisée par l'identification de régions invariables parmi les clones exprimés et par l'identification de
5 trames de lectures responsables de la production de polypeptides antigéniques et/ou pathogènes qui peuvent n'être produits que par une partie, voire un seul, des clones exprimés et dans ces conditions, l'analyse systématique des clones exprimés dans une région d'un gène
10 donné permet d'évaluer la fréquence de variation et/ou de recombinaison du génome MSRV-1 dans cette région et de définir les séquences optimales pour les applications, notamment diagnostiques; (iv) la pathologie provoquée par un rétrovirus tel que MSRV-1 peut être un effet direct de
15 son expression et des protéines ou peptides produits de ce fait, mais aussi un effet de l'activation, de l'encapsidation, de la recombinaison de génomes apparentés ou hétérologues et des protéines ou peptides produits de ces faits ; ainsi ces génomes associés à l'expression de
20 et/ou l'infection par MSRV-1 sont-ils une partie intégrante de la pathogénicité potentielle de ce virus et donc, constituent des supports de détection diagnostique et des cibles thérapeutiques particulières. De même, tout agent associé à, ou, co-facteur de ces interactions
25 responsables de la pathogénie en cause, tel que MSRV-2 ou le facteur gliotoxique décrit dans la demande de brevet publiée sous le N° FR-2 716 198, peut participer à l'élaboration d'une stratégie globale et très efficace de diagnostic, de pronostic, de suivi thérapeutique et/ou de
30 thérapeutique intégrée de la SEP notamment, mais aussi de toute autre maladie associée aux mêmes agents...

Dans ce contexte, on a fait une découverte parallèle dans une autre maladie autoimmune, la polyarthrite rhumatoïde (PR), qui a été décrite dans la
35 demande de brevet français déposée sous le N°95 02960. Cette découverte montre que, en appliquant des approches

méthodologiques similaires à celles qui furent utilisées dans les travaux de la Demanderesse sur la SEP, on a pu identifier un rétrovirus exprimé dans la PR qui partage les séquences décrites pour MSRV-1 dans la SEP et aussi, 5 la co-existence d'une séquence associée MSRV-2 également décrite dans la SEP. En ce qui concerne MSRV-1, les séquences détectées communément dans la SEP et la PR, concernent les gènes *pol* et *gag*. En l'état actuel des connaissances, on peut associer les séquences *gag* et *pol* 10 décrites aux souches MSRV-1 exprimées dans ces deux maladies.

La présente demande de brevet a pour objet différents résultats, supplémentaires par rapport à ceux déjà protégés par les demandes de brevet français :

- 15 - N° 92 04322 du 03.04.1992, publiée sous le N° 2 689 519;
- N° 92 13447 du 03.11.1992, publiée sous le N° 2 689 521;
- N° 92 13443 du 03.11.1992, publiée sous le N° 2 689 520;
- N° 94 01529 du 04.02.1994, publiée sous le N° 2 715 936;
- N° 94 01531 du 04.02.1994, publiée sous le N° 2 715 939;
- 20 - N° 94 01530 du 04.02.1994, publiée sous le N° 2 715 936;
- N° 94 01532 du 04.02.1994, publiée sous le N° 2 715 937;
- N° 94 14322 du 24.11.1994, publiée sous le N° 2 727 428;
- N° 94 15810 du 23.12.1994, publiée sous le N° 2 728 585;
- et
- 25 - la demande de brevet WO-97/06260.

La présente invention concerne tout d'abord un matériel nucléaire, qui peut consister en un matériel rétroviral, à l'état isolé ou purifié, pouvant être appréhendé ou caractérisé de différentes manières :

- 30 - il comprend une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences
 SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,
 SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,
 SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences
 35 complémentaires aux séquences (i) ; et (iii) les séquences
 équivalentes aux séquences (i) ou (ii), en particulier les

séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii);

- 5 - il code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, 10 SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137;
 - son gène pol comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 112, SEQ ID NO: 124 et leurs séquences complémentaires;
- 15 - l'extrémité 5' de son gène pol commence au nucléotide 1419 de SEQ ID NO: 130;
 - son gène pol code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % 20 d'homologie, avec la séquence peptidique SEQ ID NO: 113;
 - l'extrémité 3' de son gène gag finit au nucléotide 1418 de SEQ ID NO: 130;
 - son gène env comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le 25 groupe qui consiste en SEQ ID NO: 117, et ses séquences complémentaires;
 - son gène env comprend une séquence nucléotidique qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au nucléotide au nucléotide 233 de SEQ ID NO: 114;
- 30 - son gène env code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: °118;
 - la région U3R de son LTR 3' comprend une 35 séquence nucléotidique qui se termine au nucléotide 617 de SEQ ID NO: 114;

- la région RU5 de son LTR 5' comprend une séquence nucléotidique qui commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142;

5 - un matériel nucléique rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114;

 - le matériel nucléique rétroviral tel que défini précédemment est en particulier associé à au moins une
10 maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

L'invention concerne aussi un fragment nucléotidique qui répond au moins à l'une des définitions suivantes :

15 - il comprend ou consiste en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii)
20 les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec
25 respectivement les séquences (i) ou (ii);

 - il comprend ou consiste en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une
30 séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

D'autres objets de la présente invention sont les suivants :

35 - une sonde nucléique pour la détection d'un rétrovirus associé à la sclérose en plaques et/ou la

polyarthrite rhumatoïde, susceptible de s'hybrider spécifiquement sur tout fragment précédemment défini et appartenant au génome dudit rétrovirus; elle possède avantageusement de 10 à 100 nucléotides, de préférence de 10 à 30 nucléotides;

- une amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, qui comprend une séquence nucléotidique identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment défini précédemment, notamment une séquence nucléotidique présentant pour toute suite de 10 monomères contigus, au moins 50 %, de préférence au moins 70 % d'homologie avec au moins ladite partie dudit fragment ; de préférence la séquence nucléotidique d'une amorce de l'invention est choisie parmi SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et SEQ ID NO: 133;

- un ARN ou ADN, et notamment vecteur de répllication et/ou d'expression, comprenant un fragment génomique du matériel nucléique ou un fragment défini précédemment;

- un peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique défini précédemment, notamment un polypeptide, par exemple oligopeptide formant ou comprenant un déterminant antigénique reconnu par des sera de patients infectés par le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été réactivé; un peptide préférentiel comprend une séquence identique, partiellement ou totalement, ou équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137;

- une composition diagnostique, prophylactique, ou thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment
5 nucléotidique défini précédemment;

- un procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, comprenant les étapes consistant à mettre en contact un ARN et/ou un ADN présumé
10 appartenir ou provenant dudit rétrovirus, ou leur ARN et/ou ADN complémentaire, avec une composition comprenant un fragment nucléotidique défini ci-dessus.

Avant de détailler l'invention, différents termes utilisés dans la description et les revendications sont à
15 présent définis :

- par souche ou isolat, on entend toute fraction biologique infectante et/ou pathogène, contenant par exemple des virus et/ou des bactéries et/ou des parasites, générant un pouvoir pathogène et/ou antigénique, hébergée
20 par une culture ou un hôte vivant ; à titre d'exemple, une souche virale selon la définition précédente peut contenir un agent co-infectant, par exemple un protiste pathogène,

- le terme "MSRV" utilisé dans la présente description désigne tout agent pathogène et/ou infectant,
25 associé à la SEP, notamment une espèce virale, les souches atténuées de ladite espèce virale, ou les particules défectives interférentes ou contenant des génomes co-encapsidés ou encore des génomes recombinés avec une partie du génome MSRV-1, dérivées de cette espèce. Il est
30 connu que les virus et particulièrement les virus contenant de l'ARN ont une variabilité, consécutive notamment à des taux relativement élevés de mutation spontanée, dont il sera tenu compte ci-après pour définir la notion d'équivalence,

35 - par virus humain, on entend un virus susceptible d'infecter ou d'être hébergé par l'être humain,

- compte tenu de toutes les variations et/ou recombinaison naturelles ou induites, pouvant être rencontrées dans la pratique de la présente invention, les objets de cette dernière, définis ci-dessus et dans les
5 revendications, ont été exprimés en comprenant les équivalents ou dérivés des différents matériels biologiques définis ci-après, notamment des séquences homologues nucléotidiques ou peptidiques,

- le variant d'un virus ou d'un agent pathogène
10 et/ou infectant selon l'invention, comprend au moins un antigène reconnu par au moins un anticorps dirigé contre au moins un antigène correspondant dudit virus et/ou dudit agent pathogène et/ou infectant, et/ou un génome dont toute partie est détectée par au moins une sonde
15 d'hybridation, et/ou au moins une amorce d'amplification nucléotidique spécifique dudit virus et/ou agent pathogène et/ou infectant, dans des conditions d'hybridation déterminées bien connues de l'homme de l'art,

- selon l'invention, un fragment nucléotidique ou
20 un oligonucléotide ou un polynucléotide est un enchaînement de monomères, ou un biopolymère, caractérisé par la séquence informationnelle des acides nucléiques naturels, susceptible de s'hybrider à tout autre fragment nucléotidique dans des conditions prédéterminées,
25 l'enchaînement pouvant contenir des monomères de structures chimiques différentes et être obtenu à partir d'une molécule d'acide nucléique naturelle et/ou par recombinaison génétique et/ou par synthèse chimique ; un fragment nucléotidique peut être identique à un fragment
30 génomique du virus MSRV-1 considéré par la présente invention, notamment un gène de ce dernier, par exemple pol ou env dans le cas dudit virus ;

- ainsi un monomère peut être un nucléotide naturel d'acide nucléique, dont les éléments constitutifs
35 sont un sucre, un groupement phosphate et une base azotée; dans l'ARN le sucre est le ribose, dans l'ADN le sucre est

le désoxy-2-ribose; selon qu'il s'agit de l'ADN ou l'ARN, la base azotée est choisie parmi l'adénine, la guanine, l'uracile, la cytosine, la thymine; ou le nucléotide peut être modifié dans l'un au moins des trois éléments
5 constitutifs ; à titre d'exemple, la modification peut intervenir au niveau des bases, générant des bases modifiées telles que l'inosine, la méthyl-5-désoxycytidine, la désoxyuridine, la diméthylamino-5-désoxyuridine, la diamino-2,6-purine, la
10 bromo-5-désoxyuridine et toute autre base modifiée favorisant l'hybridation; au niveau du sucre, la modification peut consister dans le remplacement d'au moins un désoxyribose par un polyamide, et au niveau du groupement phosphate, la modification peut consister dans
15 son remplacement par des esters, notamment choisis parmi les esters de diphosphate, d'alkyl et arylphosphonate et de phosphorothioate,

- par "séquence informationnelle", on entend toute suite ordonnée de monomères, dont la nature chimique et
20 l'ordre dans un sens de référence, constituent ou non une information fonctionnelle de même qualité que celle des acides nucléiques naturels,

- par hybridation, on entend le processus au cours duquel, dans des conditions opératoires appropriées, deux
25 fragments nucléotidiques, ayant des séquences suffisamment complémentaires, s'apparient pour former une structure complexe, notamment double ou triple, de préférence sous forme d'hélice,

- une sonde comprend un fragment nucléotidique
30 synthétisé par voie chimique ou obtenu par digestion ou coupure enzymatique d'un fragment nucléotidique plus long, comprenant au moins six monomères, avantageusement de 10 à 100 monomères, de préférence 10 à 30 monomères, et possédant une spécificité d'hybridation dans des
35 conditions déterminées ; de préférence, une sonde possédant moins de 10 monomères n'est pas utilisée seule,

mais l'est en présence d'autres sondes de taille aussi courte ou non ; dans certaines conditions particulières, il peut être utile d'utiliser des sondes de taille supérieure à 100 monomères ; une sonde peut notamment être
5 utilisée à des fins de diagnostic et il s'agira par exemple de sondes de capture et/ou de détection,

- la sonde de capture peut être immobilisée sur un support solide par tout moyen approprié, c'est-à-dire directement ou indirectement, par exemple par covalence ou
10 adsorption passive,

- la sonde de détection peut être marquée au moyen d'un marqueur choisi notamment parmi les isotopes radioactifs, des enzymes notamment choisis parmi la peroxydase et la phosphatase alcaline et ceux susceptibles
15 d'hydrolyser un substrat chromogène, fluorigène ou luminescent, des composés chimiques chromophores, des composés chromogènes, fluorigènes ou luminescents, des analogues de bases nucléotidiques, et la biotine,

- les sondes utilisées à des fins de diagnostic de
20 l'invention peuvent être mises en oeuvre dans toutes les techniques d'hybridation connues, et notamment les techniques dites "DOT-BLOT", "SOUTHERN BLOT", "NORTHERN BLOT" qui est une technique identique à la technique "SOUTHERN BLOT" mais qui utilise de l'ARN comme cible, la
25 technique SANDWICH ; avantageusement, on utilise la technique SANDWICH dans la présente invention, comprenant une sonde de capture spécifique et/ou une sonde de détection spécifique, étant entendu que la sonde de capture et la sonde de détection doivent présenter une
30 séquence nucléotidique au moins partiellement différente,

- toute sonde selon la présente invention peut s'hybrider in vivo ou in vitro sur l'ARN et/ou sur l'ADN, pour bloquer les phénomènes de réplication, notamment traduction et/ou transcription, et/ou pour dégrader ledit
35 ADN et/ou ARN,

- une amorce est une sonde comprenant au moins six monomères, et avantageusement de 10 à 30 monomères, possédant une spécificité d'hybridation dans des conditions déterminées, pour l'initiation d'une polymérisation enzymatique, par exemple dans une technique d'amplification telle que la PCR (Polymerase Chain Reaction), dans un procédé d'élongation, tel que le séquençage, dans une méthode de transcription inverse ou analogue,
- 10 - deux séquences nucléotidiques ou peptidiques sont dites équivalentes ou dérivées l'une par rapport à l'autre, ou par rapport à une séquence de référence, si fonctionnellement les biopolymères correspondants peuvent jouer sensiblement le même rôle, sans être identiques,
- 15 vis-à-vis de l'application ou utilisation considérée, ou dans la technique dans laquelle elles interviennent ; sont notamment équivalentes deux séquences obtenues du fait de la variabilité naturelle, notamment mutation spontanée de l'espèce à partir de laquelle elles ont été identifiées,
- 20 ou induite, ainsi que deux séquences homologues, l'homologie étant définie ci-après,
- par "variabilité", on entend toute modification, spontanée ou induite d'une séquence, notamment par substitution, et/ou insertion, et/ou délétion de
- 25 nucléotides et/ou de fragments nucléotidiques, et/ou extension et/ou raccourcissement de la séquence à l'une au moins des extrémités; une variabilité non naturelle peut résulter des techniques de génie génétique utilisées, par exemple du choix des amorces de synthèse dégénérées ou
- 30 non, retenues pour amplifier un acide nucléique; cette variabilité peut se traduire par des modifications de toute séquence de départ, considérée comme référence, et pouvant être exprimées par un degré d'homologie par rapport à ladite séquence de référence,
- 35 - l'homologie caractérise le degré d'identité de deux fragments nucléotidiques ou peptidiques comparés ;

elle se mesure par le pourcentage d'identité qui est notamment déterminé par comparaison directe de séquences nucléotidiques ou peptidiques, par rapport à des séquences nucléotidiques ou peptidiques de référence,

5 - tout fragment nucléotidique est dit équivalent ou dérivé d'un fragment de référence, s'il présente une séquence nucléotidique équivalente à la séquence du fragment de référence ; selon la définition précédente, sont notamment équivalents à un fragment nucléotidique de
10 référence :

(a) tout fragment susceptible de s'hybrider au moins partiellement avec le complément du fragment de référence,

(b) tout fragment dont l'alignement avec le
15 fragment de référence conduit à mettre en évidence des bases contigues identiques, en nombre plus important qu'avec tout autre fragment provenant d'un autre groupe taxonomique,

(c) tout fragment résultant ou pouvant résulter de
20 la variabilité naturelle de l'espèce, à partir de laquelle il est obtenu,

(d) tout fragment pouvant résulter des techniques de génie génétique appliquées au fragment de référence,

(e) tout fragment, comportant au moins huit
25 nucléotides contigus, codant pour un peptide homologue ou identique au peptide codé par le fragment de référence,

(f) tout fragment différent du fragment de référence, par insertion, délétion, substitution d'au moins un monomère, extension, ou raccourcissement à l'une
30 au moins de ses extrémités ; par exemple, tout fragment correspondant au fragment de référence, flanqué à l'une au moins de ses extrémités par une séquence nucléotidique ne codant pas pour un polypeptide,

- par polypeptide, on entend notamment tout
35 peptide d'au moins deux acides aminés, notamment oligopeptide, protéine, extrait, séparé, ou

substantiellement isolé ou synthétisé, par l'intervention de la main de l'homme, notamment ceux obtenus par synthèse chimique, ou par expression dans un organisme recombinant,

- par polypeptide codé de manière partielle par un
5 fragment nucléotidique, on entend un polypeptide présentant au moins trois acides aminés codés par au moins neuf monomères contigus compris dans ledit fragment nucléotidique,

- un acide aminé est dit analogue à un autre acide
10 aminé, lorsque leurs caractéristiques physico-chimiques respectives, telles que polarité, hydrophobicité, et/ou basicité, et/ou acidité, et/ou neutralité, sont sensiblement les mêmes ; ainsi, une leucine est analogue à une isoleucine.

- tout polypeptide est dit équivalent ou dérivé
15 d'un polypeptide de référence, si les polypeptides comparés ont sensiblement les mêmes propriétés, et notamment les mêmes propriétés antigéniques, immunologiques, enzymologiques et/ou de reconnaissance
20 moléculaire ; est notamment équivalent à un polypeptide de référence :

(a) tout polypeptide possédant une séquence dont au moins un acide aminé a été substitué par un acide aminé analogue,

25 (b) tout polypeptide ayant une séquence peptidique équivalente, obtenue par variation naturelle ou induite dudit polypeptide de référence, et/ou du fragment nucléotidique codant pour ledit polypeptide,

(c) un mimotope dudit polypeptide de référence,

30 (d) tout polypeptide dans la séquence duquel un ou plusieurs acides aminés de la série L sont remplacés par un acide aminé de la série D, et vice versa,

(e) tout polypeptide dans la séquence duquel on a introduit une modification des chaînes latérales des
35 acides aminés, telle que par exemple une acétylation des

fonctions amines, une carboxylation des fonctions thiol, une estérification des fonctions carboxyliques,

(f) tout polypeptide dans la séquence duquel une ou des liaisons peptidiques ont été modifiées, comme par exemple les liaisons carba, rétro, inverso, rétro-inverso, réduites, et méthylène-oxy,

(g) tout polypeptide dont au moins un antigène est reconnu par anticorps dirigé contre un polypeptide de référence,

10 - le pourcentage d'identité caractérisant l'homologie de deux fragments peptidiques comparés est selon la présente invention d'au moins 50% et de préférence au moins 70 %.

Etant donné qu'un virus possédant une activité enzymatique transcriptase inverse peut être génétiquement caractérisé aussi bien sous forme d'ARN que d'ADN, il sera fait mention aussi bien de l'ADN que de l'ARN viral pour caractériser les séquences relatives à un virus possédant une telle activité transcriptase inverse, dit MSRV-1 selon la présente description.

Les expressions d'ordre utilisées dans la présente description et les revendications, telles que "première séquence nucléotidique" ne sont pas retenues pour exprimer un ordre particulier, mais pour définir plus clairement l'invention.

Par détection d'une substance ou agent, on entend ci-après aussi bien une identification, qu'une quantification, ou une séparation ou isolement de ladite substance ou dudit agent.

30 L'invention sera mieux comprise à la lecture de la description détaillée qui va suivre faite en référence aux figures annexées dans lesquelles :

La Figure 1 représente la structure générale de l'ADN proviral et l'ARN génomique de MSRV-1.

35 La Figure 2 représente la séquence nucléotidique du clone dénommé CL6-5' (SEQ ID NO: 112) et trois trames

de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 3 représente la séquence nucléotidique du clone dénommé CL6-3' (SEQ ID NO: 114) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 4 représente la séquence nucléotidique du clone dénommé C15 (SEQ ID NO: 117) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 5 représente la séquence nucléotidique du clone dénommé 5M6 (SEQ ID NO: 120) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 6 représente la séquence nucléotidique du clone dénommé CL2 (SEQ ID NO: 130) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 7 représente trois trames de lecture potentielles en acides aminés exprimées par pET28C-clone 2 et figurant sous la séquence nucléotidique.

La Figure 8 représente trois trames de lecture potentielles en acides aminés exprimées par pET21C-clone 2 et figurant sous la séquence nucléotidique.

La Figure 9 représente la séquence nucléotidique du clone dénommé LB13 (SEQ ID NO: 141) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 10 représente la séquence nucléotidique du clone dénommé LA15 (SEQ ID NO: 142) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 11 représente la séquence nucléotidique du clone dénommé LB16 (SEQ ID NO: 124) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

EXEMPLE 1 : OBTENTION D'UNE REGION CL6-5' CODANT POUR L'EXTREMITÉ N-TERMINALE DE L'INTEGRASE ET D'UNE REGION CL6-3' CONTENANT LA SEQUENCE 3' TERMINALE DU GENOME MSRV-1

5

Une 3'RACE a été effectuée sur de l'ARN total extrait de plasma d'un patient atteint de SEP. Un plasma témoin sain, traité sous les mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été

10 réalisée avec une amorce oligo dT identifiée par SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT TTT TTT TTT TTT TTT T 3') et la transcriptase inverse "ExpandTM RT" de Boehringer selon les conditions préconisées par la société. Une PCR a été effectuée avec l'enzyme Klentaq

15 (Clontech) sous les conditions suivantes : 94°C 5 min puis 93°C 1 min, 58°C 1 min, 68°C 3 min pendant 40 cycles et 68°C pendant 8 min, avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- 20 - amorce 5', identifiée par SEQ ID NO: 69
5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3' ;
- amorce 3', identifiée par SEQ ID NO: 68

Une deuxième PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la

25 région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première PCR, en utilisant 10 µl du produit d'amplification issu de la première PCR.

Amorces utilisées pour la PCR semi-nichée:

- 30 - amorce 5', identifiée par SEQ ID NO: 70
5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3' ;
- amorce 3', identifiée par SEQ ID NO: 68

Les amorces SEQ ID NO: 69 et SEQ ID NO: 70 sont spécifiques de la région pol de MSRV-1.

35 Un produit d'amplification de 1,9Kb a été obtenu pour le plasma de patient SEP. Le fragment correspondant

n'a pas été observé pour le plasma témoin sain. Ce produit d'amplification a été cloné de la façon suivante :

l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les 2 μ l de solution d'ADN ont été
5 mélangés avec 5 μ l d'eau distillée stérile, 1 μ l d'un tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2 μ l de "PCRTM VECTOR" (25 ng/ml) et 1 μ l de "T4 DNA LIGASE". Ce mélange a été incubé une nuit à 12°C. Les étapes suivantes ont été réalisées conformément aux
10 instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La
15 préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le
20 séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur Sp6 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage
25 "PRISMTM Ready Reaction AmpliTaq® FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

30 Le clone obtenu, contient une région CL6-5'codant pour l'extrémité N terminale de l'intégrase et une région CL6-3', correspondant à la région 3' terminale de MSRV-1 et permettant de définir la fin de l'enveloppe (234 pb) et les régions U3, R (401 pb) du rétrovirus MSRV1.

35 La région correspondant à l'extrémité N terminale de l'intégrase est représentée par sa séquence

nucléotidique (SEQ ID NO: 112) dans la figure 1. Les trois trames de lecture potentielles sont présentées par leur séquence aminoacide sous la séquence nucléotidique, et la séquence aminoacide de l'extrémité N-terminale de l'intégrase est identifiée par SEQID NO: 113.

La région C16-3' est représentée par sa séquence nucléotidique (SEQ ID NO: 114) dans la figure 3. Les trois trames de lecture potentielles sont présentées par leur séquence aminoacide sous la séquence nucléotidique. Une séquence aminoacide correspondant à l'extrémité C-terminale de la protéine env de MSRV-1 est identifiée par SEQ ID NO: 115.

EXEMPLE 2 : OBTENTION DU CLONE C15 CONTENANT LA REGION CODANT POUR UNE PARTIE DE L'ENVELOPPE DU RETROVIRUS MSRV-1

Une RT-PCR a été effectuée sur de l'ARN total extrait de virions concentrés par ultracentrifugation à partir d'un surnageant de culture de synoviocytes provenant d'un patient PR. La synthèse de cDNA a été réalisée avec une amorce oligo dT et la transcriptase inverse "ExpandTM RT" de Boehringer selon les conditions préconisées par la société. Une PCR a été effectuée avec l'ExpandTM Long Template PCR System (Boehringer) sous les conditions suivantes : 94°C 5 min puis 93°C 1 min, 60°C 1 min, 68°C 3 min pendant 40 cycles et 68°C pendant 8 min et avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- amorce 5', identifiée par SEQ ID NO: 69
- 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3' ;
- amorce 3', identifiée par SEQ ID NO: 116
- 5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

Une deuxième PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles

utilisées lors de la première PCR (sauf que 30 cycles ont été réalisés au lieu de 40), en utilisant 10 µl du produit d'amplification issu de la première PCR.

Amorces utilisées pour la PCR semi-nichée:

- 5 - amorce 5', identifiée par SEQ ID NO: 70
5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3' ;
- amorce 3', identifiée par SEQ ID NO: 116

Les amorces SEQ ID NO: 69 et SEQ ID NO: 70 sont spécifiques de la région pol de MSRV-1. L'amorce SEQ ID
10 NO: 116 est spécifique de la séquence FBd13 (aussi dénommé B13) et est localisée dans la région env conservée parmi les oncorétrovirus.

Un produit d'amplification de 1932 pb a été obtenu et cloné de la façon suivante :

- 15 l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été
- 20 repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides
- 25 possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur SP6 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au
- 30 séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils
- 35 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone C15 obtenu, contient une région correspondant à la région de l'enveloppe de MSRV-1, de 1481 pb.

La région env du clone C15 est représentée par sa séquence nucléotidique (SEQ ID NO: 117) dans la figure 5. Les trois trames de lecture potentielles de ce clone sont présentées par leur séquence aminoacide sous la séquence nucléotidique. La trame de lecture correspondant à une protéine env structurale MSRV-1 est identifiée par SEQ ID NO: 118.

EXEMPLE 3 : OBTENTION D'UN CLONE 5M6 CONTENANT LES SEQUENCES DE LA REGION 3' TERMINALE DE L'ENVELOPPE, SUIVIES DES SEQUENCES U3,R,U5 DE TYPE PROVIRAL MSRV-1.

Une PCR monodirectionnelle a été effectuée sur de l'ADN extrait de lymphocytes B immortalisés en culture d'un patient PR. La PCR a été effectuée avec l'Expand™ Long Template PCR System (Boehringer) sous les conditions suivantes : 94°C 3 min puis 93°C 1 min, 60°C 1 min, 68°C 3 min pendant 10 cycles , puis 93°C 1 min, 60°C 1 min avec 15 sec d'extension à chaque cycle, 68°C 3 min pendant 35 cycles et 68°C pendant 7 min et avec un volume réactionnel final de 50 µl.

L'amorce utilisée pour la PCR identifiée par SEQ ID NO: 119 est 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT AAC CG 3' ;
L'amorces SEQ ID NO: 119 est spécifique de la région env du clone C15.

Un produit d'amplification de 1673 pb a été obtenu et cloné de la façon suivante :
l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été

repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone 5M6 obtenu, contient une région correspondant à la région 3' de l'enveloppe de MSRV-1, de 492 pb suivi des régions U3, R et U5 (837 pb) de MSRV1.

Le clone 5M6 est représenté par sa séquence nucléotidique (SEQ ID NO: 120) dans la figure 7. Les trois trames de lecture potentielles de ce clone sont présentées par leur séquence aminoacide sous la séquence nucléotidique. La trame de lecture correspondant à l'extrémité C-terminale de la protéine env MSRV-1 est identifiée par SEQ ID NO: 121.

EXEMPLE 4 : OBTENTION DU CLONE LB16 CONTENANT LA REGION CODANT L'INTEGRASE DU RETROVIRUS MSRV-1

Une RT-PCR a été effectuée sur de l'ARN total traité à la DNaseI et extrait à partir d'un plexus choroïde provenant d'un patient SEP. La synthèse de cDNA a été réalisée avec une amorce oligo dT et la transcriptase inverse "Expand™ RT" de Boehringer selon les conditions préconisées par la société. Un contrôle "no RT" a été

effectué parallèlement sur le même matériel. Une PCR a été effectuée avec la Taq polymérase (Perkin Elmer) sous les conditions suivantes : 95°C 5 min puis 95°C 1 min, 55°C 1 min, 72°C 2 min pendant 35 cycles et 72°C pendant 8 min et
 5 avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- amorce 5', identifiée par SEQ ID NO: 122

5' GGC ATT GAT AGC ACC CAT CAG 3' ;

- amorce 3', identifiée par SEQ ID NO: 123

10 5' CAT GTC ACC AGG GTG GAA TAG 3'

L'amorce SEQ ID NO: 122 est spécifique de la région pol de MSRV-1 et plus précisément similaire à la région intégrase décrite précédemment. L'amorce SEQ ID NO 123 a été définie sur des séquences des clones
 15 obtenus lors d'essais préalables.

Un produit d'amplification d'environ 760 pb a été obtenu uniquement dans l'essai avec RT et a été cloné de la façon suivante :

l'ADN amplifié a été inséré dans un plasmide à l'aide du
 20 kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction
 25 des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage
 30 du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode
 35 préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™

SEP 16

Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

5 Le clone LB16 obtenu, contient les séquences correspondant à l'intégrase. La séquence nucléotidique de ce clone est identifiée par SEQ ID NO: 124 sur la figure 11, trois trames de lecture sont déterminées.

10 EXEMPLE 5: OBTENTION D'UN CLONE 2, CL2, CONTENANT EN 3' UNE PARTIE HOMOLOGUE AU GENE POL, CORRESPONDANT AU GENE PROTEASE, ET AU GENE GAG (GM3) CORRESPONDANT A LA NUCLEOCAPSIDE, ET UNE NOUVELLE REGION 5' CODANTE, CORRESPONDANT AU GENE GAG PLUS SPECIFIQUEMENT LA MATRICE
15 ET LA CAPSIDE de MSRV-1.

Une amplification par PCR a été effectuée sur de l'ARN total extrait à partir de 100 µl de plasma d'un patient atteint de SEP. Un témoin eau, traité sous les
20 mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été réalisée avec 300 pmole d'une amorce aléatoire (GIBCO-BRL, France) et la transcriptase inverse "Expand RT" (BOEHRINGER MANNHEIM, France) selon les conditions préconisées par la société. Une
25 amplification par PCR ("polymerase chain reaction") a été effectuée avec l'enzyme Taq polymerase (Perkin Elmer, France) en utilisant 10 µl de cDNA sous les conditions suivantes: 94°C 2 min, 55°C 1 min, 72°C 2 min puis 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 30 cycles et 72°C
30 pendant 7 min et avec un volume réactionnel final de 50 µl.

Amorces utilisées pour l'amplification par PCR:

- amorce 5', identifiée par SEQ ID NO: 126
5' CGG ACA TCC AAA GTG ATG GGA AAC G 3' ;
- 35 - amorce 3', identifiée par SEQ ID NO: 127
5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

Une deuxième amplification par PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions
 5 expérimentales que celles utilisées lors de la première PCR, en utilisant 10 μ l du produit d'amplification issu de la première PCR.

Amorces utilisées pour l'amplification par PCR semi-nichée:

- 10 - amorce 5', identifiée par SEQ ID NO: 128
 5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3' ;
 - amorce 3', identifiée par SEQ ID NO: 129
 5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

Les amorces SEQ ID NO: et SEQ ID NO: sont
 15 spécifiques de la région pol, clone G+E+A, plus spécifiquement la région E: position nucleotidique n° 423 à n° 448. Les amorces utilisées dans la région 5' ont été définies sur des séquences de clones obtenus lors d'essais préalables.

20 Un produit d'amplification de 1511 pb a été obtenu à partir de l'ARN extrait du plasma de patient SEP. Le fragment correspondant n'a pas été observé pour le témoin eau. Ce produit d'amplification a été cloné de la façon suivante.

25 L'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA CloningTM. Les 2 μ l de solution d'ADN ont été mélangés avec 5 μ l d'eau distillée stérile, 1 μ l d'un tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2 μ l de "PCRTM VECTOR" (25 ng/ml) et 1 μ l de "T4
 30 DNA LIGASE". Ce mélange a été incubé une nuit à 14°C. Les étapes suivantes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). Le mélange a été étalé après transformation de la ligation dans des bactéries *E. coli* INV α F'. A la fin de la procédure, les
 35 colonies blanches de bactéries recombinantes ont été repiquées pour être cultivées et permettre l'extraction

des plasmides incorporés selon la procédure dite de "minipréparation d'ADN" (17). La préparation de plasmide de chaque colonie recombinante a été coupée par l'enzyme de restriction Eco RI et analysée sur gel d'agarose. Les
 5 plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La
 10 réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISMTM Ready Reaction Amplitaq® FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les
 15 appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone obtenu, dénommé CL2, contient une région C-terminale similaire à la région 5' terminale des clones G+E+A de MSRV-1, qui permet de définir la région C-
 20 terminale du gène gag et une nouvelle région correspondante à la région N-terminale du gène gag MSRV-1.

CL2 permet de définir une région de 1511 pb présentant une phase ouverte de lecture dans la région N-terminale de 1077 pb codante pour 359 acides aminés et une
 25 phase non-ouverte de lecture, de 454 pb, correspondant à la région C-terminale du gène gag MSRV-1.

La séquence nucléotidique de CL2 est identifiée par SEQ ID NO: 130. Elle est représentée à la figure XX3,1, avec les trames de lecture potentielles en
 30 aminoacide.

Le fragment de 1077 pb de CL2 codant pour 359 acides aminés a été amplifié par PCR avec l'enzyme *Pwo* (5U/ μ l) (Boehringer Mannheim, France) en utilisant 1 μ l de la minipréparation de l'ADN du clone 2 sous les conditions
 35 suivantes : 95°C 1 min, 60°C 1 min, 72°C 2 min pendant 25

cycles et avec un volume réactionnel final de 50 μ l à l'aide des amorces:

- amorce 5' (*Bam* HI), identifié par SEQ ID NO: 132
5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer), et
 - 5 - amorce 3' (*Hind* III), identifié par SEQ ID NO: 133
5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer)
- correspondant, respectivement, à la séquence nucléotidique du clone 2 en position -9 à 21 et 1066 à 1095.

Le fragment obtenu après PCR, a été linéarisé par
10 *Bam* HI et *Hind*III et sous-cloné dans les vecteurs d'expression pET28C et pET21C (NOVAGEN) linéarisé par *Bam* HI et *Hind* III. Le séquençage de l'ADN du fragment de 1077 pb du clone 2 dans les deux vecteurs d'expression a été réalisé selon la méthode préconisée pour l'utilisation du
15 kit de séquençage "PRISMTM Ready Reaction Amplitaq® FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

20 L'expression de la séquence nucléotidique du fragment de 1077 pb du clone 2 par les vecteurs d'expression pET28C et pET21C sont identifiées par respectivement SEQ ID NO: 135 et SEQ ID NO: 137.

25 EXEMPLE 6: EXPRESSION DU CLONE 2 CHEZ *ESCHERICHIA COLI*

Les constructions pET28c-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) synthétisent, dans la souche bactérienne BL21 (DE3), une protéine en fusion N- et C-
30 terminale pour le vecteur pET28C et C-Terminale pour le vecteur pET21C avec 6 Histidines, de masse moléculaire apparente d'environ 45 kDa, mise en évidence par électrophorèse sur gel de polyacrylamide SDS-PAGE (SDS = Dodecyl Sulfate de Sodium) (Laemmli, 1970 (1)). La
35 réactivité de la protéine a été mise en évidence vis à vis

d'un anticorps monoclonal anti-Histidine (DIANOVA) par la technique de Western blot (Towbin et al., 1979 (2)).

Les protéines recombinantes pET28c-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) ont été visualisées en SDS-PAGE dans la fraction insoluble après digestion enzymatique des extraits bactériens avec 50 µl de lysozyme (10 mg/ml) et lyse par ultrasons.

Les propriétés antigéniques des antigènes recombinants pET28C-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) ont été testées par Western Blot () après solubilisation du culot bactérien avec 2% SDS et 50 mM β-mercaptoéthanol. Après incubation avec les sérums de patients atteints de sclérose en plaques, les sérums des témoins neurologiques et les sérums de témoins de centre de transfusion sanguine (CTS), les immunocomplexes ont été détectés à l'aide d'un sérum de chèvre anti-IgG et anti IgM humaines, couplé à la phosphatase alcaline.

Les résultats sont présentés dans le tableau ci-après.

TABLEAU

Réactivité de sérums atteints de sclérose en plaques et témoins avec la protéine recombinante MSRV-1 gag clone 2 (1077 pb) = pET21C-clone 2 (1077 pb) et pET28C-clone 2 (1077 pb)^a

MALADIE	NOMBRE D'INDIVIDUS TESTÉS	NOMBRE D'INDIVIDUS POSITIFS
SEP	15	6 2(+++), 2(++), 2(+)
TEMOINS		
NEUROLOGIQUES	2	1(+++)
TEMOINS		
SAINS (CTS)	22	1(+/-)

(a) Les bandelettes contenant 1,5 µg d'antigène recombinant pET-gag clone 2 (1077 pb) présentent une

réactivité contre de sérums dilués au 1/100. L'interprétation de Western Blot est basée sur la présence ou absence d'une bande pET-gag clone 2 (1077 pb) spécifique sur les bandelettes. Des contrôles positifs et négatifs sont inclus dans chaque expérience.

Ces résultats montrent que, dans les conditions techniques utilisées, environ 40% des sérums humains atteints de sclérose en plaques testés réagissent avec les protéines recombinantes pET28C-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb). Une réactivité a été observée sur un témoin neurologique et il est intéressant de noter que les ARN extraits à partir de ce sérum, après l'étape de transcriptase inverse, sont aussi amplifiés par PCR dans la région pol. Ceci suggère que des personnes n'ayant pas déclaré une SEP peuvent également héberger et exprimer ce virus. Par contre, un témoin (donneur CTS) apparemment sain, possède des anticorps anti-gag (clone 2, 1077 pb). Ce qui est compatible avec une immunité acquise contre MSR-1 en dehors d'une maladie autoimmune associée déclarée.

EXEMPLE 7: OBTENTION D'UN CLONE LB13 CONTENANT EN 3' UNE PARTIE HOMOLOGUE AU CLONE 2 CORRESPONDANT AU GENE GAG ET EN 5' UNE PARTIE HOMOLOGUE AU CLONE 5M6 CORRESPONDANT À LA RÉGION LTR U5.

Une RT-PCR ("reverse-transcriptase-polymerase chain reaction") a été effectuée à partir de l'ARN total extrait de virions, provenant de surnageants de cellules lymphocytaires B des patients atteints de sclérose en plaques, concentrés par ultracentrifugations. La synthèse de cDNA a été réalisée avec une amorce spécifique SEQ N° XXX et la transcriptase inverse "ExpandTM RT" de BOEHRINGER MANNHEIM selon les conditions préconisées par la société.

Amorce utilisée pour la synthèse du cDNA, identifiée par SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

Une amplification par PCR a été réalisée avec la
5 Tag polymérase (Perkin Elmer, France) sous les conditions
suivantes: 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 35
cycles et 72°C pendant 7 min et avec un volume réactionnel
final de 100 µl.

Amorces utilisées pour l'amplification par PCR:

10 - amorce 5', identifiée par SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- amorce 3', identifiée par SEQ ID NO: 138

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

Une deuxième amplification par PCR dite " semi-
15 nichée " a été réalisée avec une amorce 3' située à
l'intérieur de la région déjà amplifiée. Cette deuxième
amplification a été effectuée sous les mêmes conditions
expérimentales que celles utilisées lors de la première
amplification, en utilisant 10 µl du produit
20 d'amplification issu de la première PCR.

Amorces utilisées pour l'amplification par PCR " semi-
nichée ":

- amorce 5', identifiée par SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

25 - amorce 3', identifiée par SEQ ID NO: 140

5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

Les amorces SEQ ID NO: 138 et SEQ ID NO: 140 sont
spécifiques de la région gag, clone 2 position
nucléotidique n° 373-397 et n° 433-456. Les amorces
30 utilisées dans la région 5' ont été définies sur des
séquences des clones obtenus lors d'essais préalables.

Un produit d'amplification de 764 pb a été obtenu
et cloné de la façon suivante:

L'ADN amplifié a été inséré dans un plasmide à
35 l'aide du kit TA CloningTM. Les 2 µl de solution d'ADN ont
été mélangés avec 5 µl d'eau distillée stérile, 1 µl d'un

tampon de ligation 10 fois concentré "10X LIGATION
 BUFFER", 2 μ l de "pCRTM VECTOR" (25 ng/ml) et 1 μ l de "T4
 DNA LIGASE". Ce mélange a été incubé une nuit à 14°C. Les
 étapes suivantes ont été réalisées conformément au
 5 instructions du kit TA Cloning® (Invitrogen). Le mélange a
 été étalé après transformation de la ligation dans des
 bactéries *E. coli* INV α F'. A la fin de la procédure, les
 colonies blanches de bactéries recombinantes ont été
 repiquées pour être cultivées et permettre l'extraction
 10 des plasmides incorporés selon la procédure dite de
 "minipréparation d'ADN" (17). La préparation de plasmide
 de chaque colonie recombinante a été coupée par l'enzyme
 de restriction *Eco* RI et analysée sur gel d'agarose. Les
 plasmides possédant un insert détecté sous lumière UV
 15 après marquage du gel au bromure d'éthidium, ont été
 sélectionnés pour le séquençage de l'insert, après
 hybridation avec une amorce complémentaire du promoteur T7
 présent sur le plasmide de clonage du TA cloning kit®. La
 réaction préalable au séquençage a ensuite été effectuée
 20 selon la méthode préconisée pour l'utilisation du kit de
 séquençage "PRISMTM Ready Reaction Amplitaq® FS,
 DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119)
 et le séquençage automatique a été réalisé sur les
 appareils 373 A et 377 Applied Biosystems, selon les
 25 instructions du fabricant.

Le clone LB13 obtenu contient une région N-
 terminale de gène gag MSRV-1 homologue au clone 2 et un
 LTR correspondant à une partie de la région U5. Entre la
 région U5 et gag un site de fixation pour les ARN de
 30 transfert, le PBS " primer binding site " a été identifié.

La séquence nucléotidique du fragment de 764 pb
 du clone LB13 dans le plasmide "pCRTM vector" est
 représentée dans l'identificateur SEQ ID NO: 141.

Le site de fixation pour les ARN de transfert,
 35 présentant une séquence du type PBS tryptophane, a été

identifié en position nucléotidique n°342-359 du clone LB13.

Un autre clone, dénommé LA15 a été obtenu sur l'ARN total extrait de virions concentrés par
5 ultracentrifugation à partir d'un surnageant de culture de synoviocytes provenant d'un patient atteint de polyarthrite rhumatoïde. La stratégie d'amplification et clonage du clone LA15 est exactement la même qui a été utilisée pour le clone LB13.

10 La séquence nucléotidique du clone LA15 qui est représentée dans l'identificateur SEQ ID NO: 142, est très similaire au clone LB13. Ceci suggère que le rétrovirus MSVR-1 détecté dans la sclérose en plaque présente des
15 séquences similaires à celles rencontrées dans la polyarthrite rhumatoïde.

BIBLIOGRAPHIE

- (1) Laemmli U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4.
5 Nature. (1970). 227: 680-685.
- (2) Towbin H., Staehelin T. & Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some
10 applications. Proc. Natl. Acad. Sci. USA. (1979). 76: 4350-4354.

LISTE DE SEQUENCES

(2) INFORMATIONS POUR LA SEQ ID NO: 68:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR: 34 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

10 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 68:

GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT

34

(2) INFORMATIONS POUR LA SEQ ID NO: 69:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 15 (A) LONGUEUR: 30 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

20 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 69:

GCCATCAAGC CACCCAAGAA CTCTTAACTT

30

(2) INFORMATIONS POUR LA SEQ ID NO: 70:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 25 (A) LONGUEUR: 30 paires de bases
(B) TYPE: nucléotide
(C) NOMBRE DE BRINS: simple
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 70:

CCAATAGCCA GACCATTATA TACTACTAATT

30

(2) INFORMATIONS POUR LA SEQ ID NO: 112

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 310 paires de bases
 (B) TYPE: nucléotidique
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 112:

GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC CCCAGTACTC 60
 AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT 120
 10 AGCCACTGAG GAAGGAAAAA TACTTTTACC TGCAGCTAAC CAACAGAAAT TACTTAAAC 180
 CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT 240
 TACTGGACCA GGCCTTTTCA AAACCTATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300
 AAGAAATAAT 310

15 (2) INFORMATIONS POUR LA SEQ ID NO: 113

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 103 acides aminés
 (B) TYPE: peptidique
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 113:

LIEGPLVWGNPLWETKPQYSAGKIEXETSQGHTFLPSRWLATEEGKILSPAANQQKLLKTLHQTFHLGID
 STHQMAKLLFTGPGLFKTIKKIVRGCEVCQRNN

25 (2) INFORMATIONS POUR LA SEQ ID NO: 114

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 635 paires de bases
 (B) TYPE: nucléotidique
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 114:

CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC TGTAATACTA 60
 CAAATTGTTT TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC 120
 CTGGACCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG 180
 35 GAAATCTCAA CTGCACAACC CCTACTATGC CCCAATTCAG CGGGAAGCAG TTAGAGCGGT 240
 CATCAGCCAA CCTCCCCAAC AGCACTTGGG TTTCTCTGTT GAGAGGGGGG ACTGAGAGAC 300

AGGACTAGCT GGATTTCTTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT 360
 GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCCGACCAAT CAGAGAGCTC 420
 ACTAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG 480
 AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAAACCCA GGCATTCGAG CCGGCAACGG 540
 5 CAACCCCCTT TGGGTCCCCT CCCTTTGTAT GGGCGCTCTG TTTTCACTCT ATTTCACTCT 600
 ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

(2) INFORMATIONS POUR LA SEQ ID NO: 115

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 10 (A) LONGUEUR: 77 acides aminés
 (B) TYPE: peptidique
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 115:

15 PCIFNLLVKFVSSRIKTVKLQIVLQMEHQMESMTKIHRGPLDRPASPCSDVNDIEGTPPEEISTAQPLLC
 PNSAGSS

(2) INFORMATIONS POUR LA SEQ ID NO: 116

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 20 (A) LONGUEUR: 32 paires de bases
 (B) TYPE: nucléotidique
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 116:

25 TGGGGTTCCA TTTGTAAGAC CATCTGTAGC TT 32

(2) INFORMATIONS POUR LA SEQ ID NO: 117

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR: 1481 paires de base
 (B) TYPE: nucléotidique
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 117:

ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60
 35 GCACCCCTC CATGCTGCTG TACAACCACT AGCTCCCCTT ACCAAGAGTT TCTATGAAGA 120
 ACGCGGCTTC CTGGAAATAT TGATGCCCCA TCATATAGGA GTTATCTAA GGGAACTCC 180

ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG 240
 CATGCAAATA CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 300
 GGAGCCACTG TCTGTTGGAC TTA CTTCACC CATAACAGTA TGTCTGATGG GGGTGGAATT 360
 CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420
 5 AGCACCCCTA GCCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480
 CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540
 CAAAACCCTA CTAAGTGTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTC AATC 600
 CCTGTTCTTG AACAATGGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 660
 GGACCTCTTG TTTCCAATCT GGAAATAACC CATACTCAA ACCTCACCTG TGTA AAAATTT 720
 10 AGCAATACTA TAGACACAAC CAGCTCCCAA TGCATCAGGT GGGTAACACC TCCACACGA 780
 ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG 840
 AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC 900
 ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960
 CTTCCTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020
 15 ACAACCTCTA CTCAGTTCTA CTACAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080
 GTCAGTACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 1140
 CAAAATCGAA GAGCTTTAGA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTTA 1200
 GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260
 ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAA ACACCGAACG CTGGGGCCTC 1320
 20 CTCAGCCAAT GGATGCCCTG GGTCTCCTC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380
 TTA CTCTCT TGGACCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440
 GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

(2) INFORMATIONS POUR LA SEQ ID NO: 118

25 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 493 acides aminés

(B) TYPE: peptidique

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 118:

MALPYHTFLFTVLLPPFALTAPPPCCCTTSSSPYQEFLEXRTRLPGNIDAPSYRSLSKGNSTFTAHTHMPR
 NCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTFYFHTSMSDGGGIQQQAREKQVKEAISQLTRGH
 STSPYKGLVLSKLHETLRTHRLVSLFNNTLTRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFS
 TEINTTSVLVGPLVSNLEITHTSNLTVCVKFSNTIDTSSQCIWVTPPTRIVCLPSGIFFVCGTSAYHCL
 35 NGSSSESMCFLSFLVPPMTIYTEQDLNHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYFKL
 SQEINGDMEQVTDLSVTLQDQLNSLAAVVLQNRRLDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE

IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMVLO
MEP

(2) INFORMATIONS POUR LA SEQ ID NO: 119

5 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 32 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

10 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 119:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG

32

(2) INFORMATIONS POUR LA SEQ ID NO: 120

(i) CARACTERISTIQUES DE LA SEQUENCE:

15 (A) LONGUEUR: 1329 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 120:

20	TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT GTTTATTTTT	60
	AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA	120
	AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT	180
	CCTCAGCCAA TGGATGCCCT GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT	240
	TTTACTCCTC TTTGGACCCT GTATCTTCAA CTTCTTGTT AAGTTTGTCT CTTCCAGAAT	300
25	TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT CCATGACTAA	360
	AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA	420
	AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG	480
	AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA	540
	GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT	600
30	ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG	660
	ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA CAGGAGGTAA AGCAATAGCC	720
	AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA	780
	TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT	840
	CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG	900
35	CTCAAGCTGA GCTTTTGTTT GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT	960
	GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT	1020

ACCCATG GCC ACTCCCGATC AGGCTAAAGG CTTGCCATTG TTCCTGCATG GCTAAGTGCC 1080
 TGGGTTTGTC CTAATAGAAC TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC 1140
 CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGGTA 1200
 TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TGCCACCATT TGGGAAGTGG 1260
 5 CCCACTGCCA TTTTGGTAGC GGCCACCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320
 CCAGTAACA 1329

(2) INFORMATIONS POUR LA SEQ ID NO: 121

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 10 (A) LONGUEUR: 162 acides aminés
 (B) TYPE: peptide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 121:

15 QNRRALDLLTAKRGGTCLFLGEECCXYVNQSGIITEKVKEIXDRIXCRAEDLQNTAPWGLLSQWMPWTLF
 FLGPLAAIIFLLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEPQMOSMTKIYRGPLDRPARLCSDVNDIE
 VTPPEEISTAQPLLHSNSVGSS

(2) INFORMATIONS POUR LA SEQ ID NO: 122

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 20 (A) LONGUEUR: paires de base
 (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 122:

25 GGCATTGATA GCACCCATCA G 21

(2) INFORMATIONS POUR LA SEQ ID NO: 123

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR: 21 paires de base
 (B) TYPE: nucléotide
 (C) NOMBRE DE BRINS: simple
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 123:

35 CATGTCACCA GGGTGAATA G 21

(2) INFORMATIONS POUR LA SEQ ID NO: 124

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 124:

10

15

(2) INFORMATIONS POUR LA SEQ ID NO: 126

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 21 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 126:

CATGTCACCA GGGTGAATA G

21

25

(2) INFORMATIONS POUR LA SEQ ID NO: 127

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 26 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 127:

GGACAGGAAA GTAAGACTGA GAAGGC

26

35 (2) INFORMATIONS POUR LA SEQ ID NO: 128

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

5 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 128:
cf exemple 5

(2) INFORMATIONS POUR LA SEQ ID NO: 129

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 10 (A) LONGUEUR: paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

15 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 129:
cf exemple 5

(2) INFORMATIONS POUR LA SEQ ID NO: 130

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 20 (A) LONGUEUR: 1511 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 130:

	CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG	60
25	ATTTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG	120
	GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA	180
	GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACCTC	240
	ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA	300
	CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC	360
30	GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCTCCG	420
	ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTCCGC CCAGCCAGAG TGCCTGTACC	480
	TTTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAATTCT CAGATAACCC	540
	TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA	600
	TATAATGTTA CTAATAATC AGACACTAAC CCCAATGAG AGAAGTGCCG CTGTAACCTG	660
35	AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA	720
	GGAAAGAACA ACTCCACAG GCCAGCAGGC AGTTCCAGT GTAGACCCTC ATTGGGACAC	780

AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTTGCTA ACTTGCGTGC TAGAAGGACT 840
 GAGGAAAAC AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900
 GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960
 CCTGTACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020
 5 AGCTGCAGAC ATTAGAAAAA ACTTCAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080
 ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140
 CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTAGTCATG GCCCTCAGGC 1200
 AAGCAGACTT TGGAGGCTCT GCAAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260
 CTGGCTTCCA GTGCGGTCTA CAAGGACACT TTAAAAAGA TTATCCAAGT AGAAATAAGC 1320
 10 CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380
 GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440
 GCGCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCGCGGGTA TGTTTGACCA 1500
 TTGAGAGCCA A 1511

15 (2) INFORMATIONS POUR LA SEQ ID NO: 131

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 347 acides aminés
- (B) TYPE: peptide
- (C) NOMBRE DE BRINS: simple
- 20 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 131:

LERILENWDQCDTQTLRKKRFIFFCSTAWPQYPLQGRETWLPEGSINYNIILQLDLFCRKEGKWSEVPYV
 QTFPSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPTPSSTNKDPPLTQTVQKEIDKGVNNEPKSANIPR
 LCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLGKFSNDPDGYIDVLQGLGQSFDLTWRDIMLLNQTILT
 25 PNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTGQQAVPSVDPHWDTESEHGDWCHKHLLTCVLEGL
 RKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRKHTSLSPDSIEGQLILKDKFITQSAADIRKNFKS
 LP

(2) INFORMATIONS POUR LA SEQ ID NO: 132

30 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 30 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

35 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 132:

TGCTGGAATT CGGGATCCTA GAACGTATTC

(2) INFORMATIONS POUR LA SEQ ID NO: 133

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 30 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 133:

AGTTCTGCTC CGAAGCTTAG GCAGACTTTT

30

(2) INFORMATIONS POUR LA SEQ ID NO: 135

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: acides aminés
- (B) TYPE: peptide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 135:

MGSSHHHHHSSGLVPRGSHMASMTGGQQMGRILERILENWDQCDTQTLRKKRFIFFCSTAWPQYPLQGR
 ETWLPEGSINYNIIQLDLFCRKEGKWSEVPYVQTFSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPT
 PSSTNKDPPLTQTVQKEIDKGVNNEPKSANIPRLCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLGKFS
 DNPBGYIDVLQGLGQSFDLTWRDIMLLNQTLPNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTG
 QQAVPSVDPHWDTESEHGDWCHKHLLTCVLEGLRKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRK
 HTSLSPDSIEGQLILKDKFITQSAADIRKNFKSLPKLAAALEHHHHHH

10 (2) INFORMATIONS POUR LA SEQ ID NO: 137

(i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: acides aminés
- (B) TYPE: peptide
- (C) NOMBRE DE BRINS: simple
- 15 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 137:

MASMTGGQQMGRILERILENWDQCDTQTLRKKRFFCSTAWPQYPLQGRETWLPEGSINYNIILQLDLF
 CRKEGKWSEVPYVQTFSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPTPSSTNKDPPLTQTVQKEIDK
 GVNNEPKSANIPRLCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLGKFSDNPDGYIDVLQGLGQSFDLT
 20 WRDIMLLLNQTLTPNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTGQQAVPSVDPHWDTESEHGDW
 CHKHLITCVLEGLRKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRKHTSLSPDSIEGQLILKDKFI
 TQSAADIRKNFKSLPKLAAALEHHHHHH

(2) INFORMATIONS POUR LA SEQ ID NO: 138

25 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 25 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 138:

CTTGGAGGGT GCATAACCAG GGAAT

25

(2) INFORMATIONS POUR LA SEQ ID NO: 139

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 35 (A) LONGUEUR: 20 paires de base
- (B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 139:

TGTCCGCTGT GCTCCTGATC

20

5

(2) INFORMATIONS POUR LA SEQ ID NO: 140

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 25 paires de base

(B) TYPE: nucléotide

10

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 140:

CTATGTCCTT TTGGACTGTT TGGGT

25

15 (2) INFORMATIONS POUR LA SEQ ID NO: 141

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 764 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

20

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 141:

TGTCCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTGCC	TCTCCCAATT	GGGCTAAAGG	60
CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTTCATC	CTAATCGAGC	TGAACACTAG	120
TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	CCATGGCTTC	TAATAGAGCT	ATAAACTCA	180
25 CTGCATGGTC	CAAGATTCCA	TTCCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	240
ACACAAGGCT	TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTTGGAAGC	AGCCCGCCAC	300
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	AGGTAACAAT	TTGGTGACCA	CGAAGGGACC	360
TGAATCCGCA	ACCATGAAGG	GATCTCCAAA	GCAATTGGAA	ATGTTCTCTC	CAAGGCAAAA	420
ATGCCCTTAA	GATGTATTCT	GGAGAATTGG	GACCAATTTG	ACCCTCAGAC	AGTAAGAAAA	480
30 AAATGACTTA	TATTCTTCTG	CAGTACCGCC	CTGGCCACGA	TATCCTCTTC	AAGGGGGAGA	540
AACCTGGCCT	CCTGAGGGAA	GTATAAATTA	TAACACCATC	TTACAGCTAG	ACCTGTTTTG	600
TAGAAAAGGA	GGCAAATGGA	GTGAAGTGCC	ATATTTACAA	ACTTCTTTT	CATTAAAAGA	660
CAACTCGCAA	TTATGTTAAC	AGTGTGATTT	GTGTTCTTAC	ACGGAAGCCC	TCAGATTCTA	720
CTCCCCACCC	CCGGCATCTC	CCCTGAATCC	CTCCCCAACT	TATT		764

35

(2) INFORMATIONS POUR LA SEQ ID NO: 142

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 800 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

5

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 142:

	TGTCCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTGCC	TCTCCCAATT	GGGCTAAAGG	60
	CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTCATC	CTAATCGAGC	TGAACACTAG	120
	TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	CCATGGCTTC	TAATAGAGCT	ATAACACTCA	180
10	CTGTCATGGTC	CAAGATTCCA	TTCCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	240
	ACACAAGGCT	TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTGGAAGC	GGCCCCGCCAC	300
	TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	CAGGTAACAA	TTTGGTGACC	ACGAAGGGAC	360
	CTGAATCCGC	AACCATGAAG	GGATCTCCAA	AGCAATTGGA	AATGTTTCCTC	CCAAGGCAAA	420
	AATGCCCCTA	AGATGTATTC	TGGAGAATTG	GGACCAATCT	GACCCCTCAGA	CAGTAAGAAA	480
15	AAAAATGACT	TATATTCTTC	TGCAGTACCG	CCTGGCCACG	GATATCCTCT	TCAAGGGGGA	540
	GAAACCTGGC	CTCCTGAGGG	AAGTATAAAT	TATAACACCA	TCTTACAGCT	AGACCTGTTT	600
	TGTAGAAAAG	GAGGCAAATG	GAGTGAAGTG	CCATATTTAC	AAACTTTCTT	TTCATTAAAA	660
	GACAACTCGC	AATTATGTAA	ACAGTGTGAT	TTGTGTCCTA	CAGGAAGCCC	TCAGATCTAC	720
	CTCCCTACCC	CGGCATCTCC	CTGACTCCTT	CCCCAACTAA	TAAGGACCCA	CTTCAGCCCA	780
20	AACAGTCCAA	AAGGACATAG					800

REVENDICATIONS

1. Matériel nucléique, à l'état isolé ou purifié, comprenant une séquence nucléotidique choisie dans le
5 groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux
10 séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii).

2. Matériel nucléique, à l'état isolé ou purifié,
15 codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118,
20 SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

3. Matériel nucléique rétroviral, dont le gène pol comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 112, SEQ ID NO: 124 et leurs
25 séquences complémentaires.

4. Matériel nucléique rétroviral, dont l'extrémité 5' du gène pol commence au nucléotide 1419 de SEQ ID NO: 130.

5. Matériel nucléique rétroviral, dont le gène pol
30 code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence peptidique SEQ ID NO: 113.

6. Matériel nucléique rétroviral, dont l'extrémité
35 3' du gène gag finit au nucléotide 1418 de SEQ ID NO: 130.

7. Matériel nucléique rétroviral, dont le gène env comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 117, et ses séquences
5 complémentaires.

8. Matériel nucléique rétroviral, dont le gène env comprend une séquence nucléotidique qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au nucléotide au nucléotide 233 de SEQ ID NO: 114.

10 9. Matériel nucléique rétroviral, dont le gène env code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: 118.

15 10. Matériel nucléique rétroviral dont la région U3R du LTR 3' comprend une séquence nucléotidique qui se termine au nucléotide 617 de SEQ ID NO: 114.

11. Matériel nucléique rétroviral dont la région RU5 du LTR 5' comprend une séquence nucléotidique qui
20 commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142.

12. Matériel nucléique rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114.

25 13. Matériel nucléique rétroviral selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il est associé à au moins une maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

30 14. Fragment nucléotidique comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii)
35 les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en

particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii).

5 15. Fragment nucléotidique selon la revendication 14, consistant en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences
 SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,
 SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,
 10 SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins
 15 70 % d'homologie avec respectivement les séquences (i) ou (ii).

16. Fragment nucléotidique comprenant une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins
 20 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en
 SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118,
 SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

17. Fragment nucléotidique selon la revendication
 25 16, consistant en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en
 30 SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121,
 SEQ ID NO: 135 et SEQ ID NO: 137.

18. Sonde nucléique pour la détection d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle est
 35 susceptible de s'hybrider spécifiquement sur tout fragment

selon l'une quelconque des revendications 14 à 17, appartenant au génome dudit rétrovirus.

19. Sonde selon la revendication 18, caractérisée en ce qu'elle possède de 10 à 100 nucléotides, de
5 préférence de 10 à 30 nucléotides.

20. Amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle comprend une séquence nucléotidique
10 identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment selon l'une quelconque des revendications 8 à 11, notamment une séquence nucléotidique présentant pour toute suite de 10 monomères contigus, au moins 50 %, de préférence au moins
15 70 % d'homologie avec au moins ladite partie dudit fragment.

21. Amorce selon la revendication 20, caractérisée en ce que sa séquence nucléotidique est choisie parmi
SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122,
20 SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,
SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et
SEQ ID NO: 133.

22. ARN ou ADN, et notamment vecteur de répllication et/ou d'expression, comprenant un fragment
25 génomique du matériel nucléique selon l'une quelconque des revendications 1 à 7 ou un fragment selon l'une quelconque des revendications 14 à 17.

23. Peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique selon l'une
30 quelconque des revendications 14 à 17, notamment polypeptide, par exemple oligopeptide formant ou comprenant un déterminant antigénique reconnu par des sera de patients infectés par le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été réactivé.

35 24. Peptide selon la revendication 23 comprenant une séquence identique, partiellement ou totalement, ou

équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

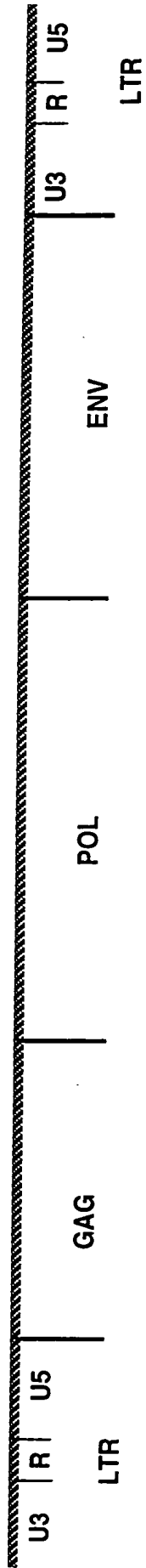
25. Composition diagnostique, prophylactique, ou
5 thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.
- 10 26. Procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, caractérisé en ce que l'on met en contact un ARN et/ou un ADN présumé appartenir ou
provenant dudit rétrovirus, ou leur ARN et/ou ADN
15 complémentaire, avec une composition comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.

ABREGE

MATERIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES NOTAMMENT ASSOCIES A LA SCLEROSE EN PLAQUES ET/OU LA POLYARTHRITE RHUMATOIDE, A DES FINS DE DIAGNOSTIC, PROPHYLACTIQUES ET THERAPEUTIQUES

Matériel nucléique, à l'état isolé ou purifié, et fragment nucléotidique, comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii), et utilisations pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde.

ADN PROVIRAL



ARN GENOMIQUE (VIRION)

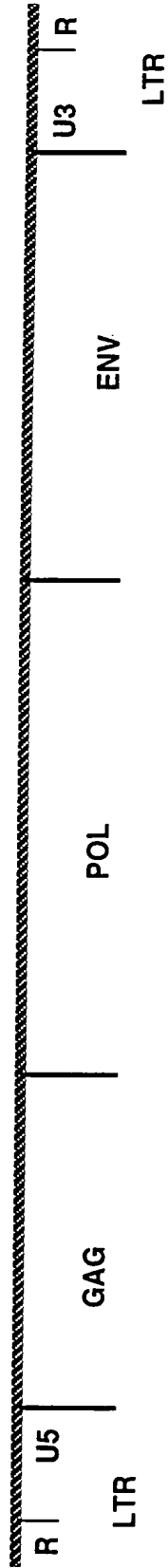


FIG. 1

FIG. 2

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCTTATAGAA	GGACCCCTAG	TATGGGGTAA	TCCCCTCTGG	GAAACCAAGC	50
A Y R R	T P S	M G .	S P L G	N Q A	
L I E	G P L V	W G N	P L W	E T K P	
L . K	D P .	Y G V I	P S G	K P S	
CCAGTACTC	AGCAGGAAAA	ATAGAATAGG	AAACCTCACA	AGGACATACT	100
P V L	S R K N	R I G	N L T	R T Y F	
Q Y S	A G K	I E .	E T S Q	G H T	
P S T Q	Q E K	. N R	K P H K	D I L	
TTCTCCCT	CCAGATGGCT	AGCCACTGAG	GAAGGAAAAA	TACTTTTACC	150
P P L	Q M A	S H .	G R K N	T F T	
F L P S	R W L	A T E	E G K I	L S P	
S S P	P D G .	P L R	K E K	Y F H L	
TGCAGCTAAC	CAACAGAAAT	TACTTAAAC	CCTTCACCAA	ACCTTCCACT	200
C S .	P T E I	T . N	P S P N	L P L	
A A N	Q Q K L	L K T	L H Q	T F H L	
Q L T	N R N	Y L K P	F T K	P S T	
TAGGCATTGA	TAGCACCCAT	CAGATGGCCA	AATTATTATT	TACTGGACCA	250
R H .	. H P S	D G Q	I I I	Y W T R	
G I D	S T H	Q M A K	L L F	T G P	
. A L I	A P I	R W P	N Y Y L	L D Q	
GGCCTTTTCA	AAACTATCAA	GAAGATAGTC	AGGGGCTGTG	AAGTGTGCCA	300
P F Q	N Y Q	E D S Q	G L .	S V P	
G L F K	T I K	K I V	R G C E	V C Q	
A F S	K L S R	R . S	G A V	K C A K	
AAGAAATAAT					310
K K .					
R N N					
E I					

FIG 2 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCCTGTATCT	TTAACCTCCT	TGTTAAGTTT	GTCTCTTCCA	GAATCAAAAC	50
P C I F	N L L	V K F	V S S R	I K T	
P V S	L T S L	L S L	S L P	E S K L	
L Y L	. P P	C . V C	L F Q	N Q N	
TGTAATACTA	CAAATTGTTC	TTCAAATGGA	GCACCAGATG	GAGTCCATGA	100
V K L	Q I V L	Q M E	H Q M	E S M T	
. N Y	K L F	F K W S	T R W	S P .	
C K T T	N C S	S N G	A P D G	V H D	
CTAAGATCCA	CCGTGGACCC	CTGGACCGGC	CTGCTAGCCC	ATGCTCCGAT	150
K I H	R G P	L D R P	A S P	C S D	
L R S T	V D P	W T G	L L A H	A P M	
. D P	P W T P	G P A	C . P	M L R C	
GTTAATGACA	TTGAAGGCAC	CCCTCCCGAG	GAATCTCAA	CTGCACAACC	200
V N D I	E G T	P P E	E I S T	A Q P	
L M T	L K A P	L P R	K S Q	L H N P	
. . H	. R H	P S R G	N L N	C T T	
CCTACTATGC	CCCAATTCAG	CGGGAAGCAG	TTAGAGCGGT	CATCAGCCAA	250
L L C	P N S A	G S S	. S G	H Q P T	
Y Y A	P I Q	R E A V	R A V	I S Q	
P T M P	Q F S	G K Q	L E R S	S A N	
CCTCCCCAAC	AGCACTTGGG	TTTTCTGT	GAGAGGGGGG	ACTGAGAGAC	300
S P T	A L G	F S C	. E G G	L R D	
P P Q Q	H L G	F P V	E R G D	. E T	
L P N	S T W V	F L L	R G G	T E R Q	
AGGACTAGCT	GGATTTCCTA	GGCCAACGAA	GAATCCCTAA	GCCTAGCTGG	350
R T S W	I S .	A N E	E S L S	L A G	
G L A	G F P R	P T K	N P .	A . L G	
D . L	D F L	G Q R R	I P K	P S W	

4/27

FIG 3

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GAAGGTGACT	GCATCCACCT	CTAAACATGG	GGCTTGCAAC	TTAGCTCACA	400
K V T A S T S	K H G A C N	L A H T			
R . L H P P	L N M G L A T	. L T			
E G D C I H L	. T W G L Q L	S S H			
CCCGACCAAT	CAGAGAGCTC	ACTAAAATGC	TAATTAGGCA	AAAATAGGAG	450
R P I R E L	T K M L I R Q	K . E			
P D Q S E S S	L K C . L G K	N R R			
P T N Q R A H	. N A N . A	K I G G			
GTAAAGAAAT	AGCCAATCAT	CTATTGCGCTG	AGAGCACAGC	GGGAGGGACA	500
V K K . P I I	Y C L R A Q R	E G Q			
. R N S Q S S	I A . E H S	G R D K			
K E I A N H	L L P E S T A	G G T			
AGGATCGGGA	TATAAACCCA	GGCATTCGAG	COGGCAACGG	CAACCCCTT	550
G S G Y K P R	H S S R Q R	Q P P L			
D R D I N P	G I R A G N G	N P L			
R I G I . T Q	A F E P A T A	T P F			
TGGGTCCCTT	CCCTTGTGAT	GGCGCTCTG	TTTTCACTCT	ATTTCACTCT	600
G P L P L Y	G R S V F T L	F H S			
W V P S L C M	G A L F S L Y	F T L			
G S P P F V W	A L C F H S	I S L Y			
ATTAAATCTT	GCAACTGAAA	AAAAAAAAAA	AAAAA		635
I K S C N . K	K K K K	K			
L N L A T E K	K K K K	K			
. I L Q L K	K K K K	K			

5/27

FIG 4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCTCC	CTTATCATA	CTTTCTCTTT	ACTGTTCTCT	TACCCCTTT	50
M A L P	Y H T	F L F	T V L L	P P F	
W P S	L I I L	F S L	L F S	Y P L S	
G P P	L S Y	F S L Y	C S L	T P F	
CGCTCTCACT	GCACCCCTC	CATGCTGCTG	TACAACCAGT	AGCTCCCTT	100
A L T	A P P P	C C C	T T S	S S P Y	
L S L	H P L	H A A V	Q P V	A P L	
R S H C	T P S	M L L	Y N Q	. L P L	
ACCAAGAGTT	TCTATGAAGA	ACGCGGCTTC	CTGGAAATAT	TGATGCCCCA	150
Q E F	L . R	T R L P	G N I	D A P	
T K S F	Y E E	R G F	L E I L	M P H	
P R V	S M K N	A A S	W K Y	. C P I	
TCATATAGGA	GTTTATCTAA	GGGAACTCC	ACCTTCACTG	CCCACACCCA	200
S Y R S	L S K	G N S	T F T A	H T H	
H I G	V Y L R	E T P	P S L	P T P I	
I . E	F I .	G K L H	L H C	P H P	
TATGCCCCGC	AACTGCTATA	ACTCTGCCAC	TCTTTGCATG	CATGCAAATA	250
M P R	N C Y N	S A T	L C M	H A N T	
C P A	T A I	T L P L	F A C	M Q I	
Y A P Q	L L .	L C H	S L H A	C K Y	
CTCATTATTG	GACAGGGAAA	ATGATTAATC	CTAGTTGTCC	TGGAGGACTT	300
H Y W	T G K	M I N P	S C P	G G L	
L I I G	Q G K	. L I	L V V L	E D L	
S L L	D R E N	D . S	. L S	W R T W	
GGAGCCACTG	TCTGTTGGAC	TTACTTCACC	CATACCAGTA	TGTTCTGATGG	350
G A T V	C W T	Y F T	H T S M	S D G	
E P L	S V G L	T S P	I P V	C L M G	
S H C	L L D	L L H P	Y Q Y	V . W	

6/27

FIG 4 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ACCTCACCTG	TGTAAAATTT	AGCAATACTA	TAGACACAAC	CAGCTCCCAA	750
L T C V K F	S N T I	D T T	S S Q		
T S P V .	N L A I L	. T Q P	A P N		
P H L C K I .	Q Y Y	R H N	Q L P M		
TGCATCAGGT	GGGTAAACACC	TCCCACACGA	ATAGTCTGCC	TACCCCTCAGG	800
C I R W V T P	P T R I V C L	P S G			
A S G G .	H L P H E	. S A Y P Q E			
H Q V G N T	S H T N	S L P T L R			
AATATTTTTT	GTCTGIGGTA	CCTCAGCCTA	TCATTGTTTG	AATGGCTCTT	850
I F F V C G T	S A Y H C L	N G S S			
Y F L S V V	P Q P I I V .	M A L			
N I F C L W Y	L S L S L F E	W L F			
CAGAATCTAT	GTGCTTCCTC	TCATTCTTAG	TGCCCCCTAT	GACCATCTAC	900
E S M C F L	S F L V P P M	T I Y			
Q N L C A S S	H S . C P L .	P S T			
R I Y V L P L	I L S A P Y	D H L H			
ACTGAACAAG	ATTATATAAA	TCATGTGCTA	CCTAAGCCCC	ACAACAAAAG	950
T E Q D L Y N	H V V P K P H	N K R			
L N K I Y T I	M S Y L S P T T K E				
. T R F I Q	S C R T . A P	Q Q K			
AGTACCCATT	CTTCCTTTTG	TTATCAGAGC	AGGAGTGCTA	GCCAGACTAG	1000
V P I L P F V	I R A G V L	G R L G			
Y P F F L L	L S E Q E C .	A D .			
S T H S S F C	Y Q S R S A R	Q T R			
GTACTGGCAT	TGGCAGTATC	ACAACCTCTA	CTCAGTTCTA	CTACAAACTA	1050
T G I G S I	T T S T Q F Y	Y K L			
V L A L A V S	Q P L L S S T	T N Y			
Y W H W Q Y H	N L Y S V L	L Q T I			

7/27

FIG 4 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCTCAAGAAA	TAAATGGTGA	CATGGAACAG	GTCACTGACT	CCCTGGTCAC	1100
S Q E I	N G D	M E Q	V T D S	L V T	
L K K	. M V T	W N R	S L T	P W S P	
S R N	K W .	H G T G	H . L	P G H .	
CTTGCAAGAT	CAACTTAACT	CCCTAGCAGC	AGTAGTCCTT	CAAAATCGAA	1150
L Q D	Q L N S	L A A	V V L	Q N R R	
C K I	N L T	P . Q Q	. S F	K I E	
L A R S	T . L	P S S	S S P S	K S K	
GAGCTTTAGA	CTTGCTAACC	GCCAAAAGAG	GGGGAACCTG	TTTATTTTTTA	1200
A L D	L L T	A K R G	G T C	L F L	
E L .	T C .	P P K E	G E P V	Y F .	
S F R	L A N R	Q K R	G N L	F I F R	
GGAGAAGAAC	GCTGTTATTA	TGTTAATCAA	TCCAGAATTG	TCACTGAGAA	1250
G E E R	C Y Y	V N Q	S R I V	T E K	
E K N	A V I M	L I N	P E L	S L R K	
R R T	L L L	C . S I	Q N C	H . E	
AGTTAAAGAA	ATTCGAGATC	GAATACAATG	TAGAGCAGAG	GAGCTTCAAA	1300
V K E	I R D R	I Q C	R A E	E L Q N	
L K K	F E I	E Y N V	E Q R	S F K	
S . R N	S R S	N T M	. S R G	A S K	
ACACCGAACG	CTGGGGCCTC	CTCAGCCAAT	GGATGCCCTG	GGTTCTCCCC	1350
T E R	W G L	L S Q W	M P W	V L P	
T P N A	G A S	S A N	G C P G	F S P	
H R T	L G P P	Q P M	D A L	G S P L	
TTCTTAGGAC	CTCTAGCAGC	TCTAATATTG	TTACTCCTCT	TTGGACCCCTG	1400
F L G P	L A A	L I L	L L L F	G P C	
S . D	L . Q L	. Y C	Y S S	L D P V	
L R T	S S S	S N I V	T P L	W T L .	

8/27

FIG 4 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TATCTTTAAC	CTCCTTGTTA	AGTTTGTC	TTCCAGAATT	GAAGCTGTAA	1450
I F N	L L V K	F V S	S R I	E A V K	
S L T	S L L	S L S L	P E L	K L .	
Y L .	P P C .	V C L	F Q N .	S C K	
AGCTACAGAT GGTCTTACAA ATGGAACCCC A					1481
L Q M	V L Q	M E P			
S Y R W	S Y K	W N P			
A T D	G L T N	G T P			

9/27

FIG. 5

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCAAATCGA	AGAGCTTTAG	ACTTGCTAAC	CGCCAAAAGA	GGGGGAACCT	50
S K S K	S F R	L A N	R Q K R	G N L	
Q N R	R A L D	L L T	A K R	G G T C	
K I E	E L .	T C .	P P K E	G E P	
GTTTATTTTT	AGGGGAAGAA	TGCTGTTAGT	ATGTTAATCA	ATCTGGAATC	100
F I F	R G R M	L L V	C . S	I W N H	
L F L	G E E	C C .	Y V N Q	S G I	
V Y F .	G K N	A V S	M L I N	L E S	
ATTACTGAGA	AAGTTAAAGA	AATTTCAGAT	CGAATATAAT	GTACAGCAGA	150
Y . E	S . R	N L R S	N I M .	S R	
I T E K	V K E	I . D	R I . C	R A E	
L L R	K L K K	F E I	E Y N	V E Q R	
GGACCTTCAA	AACACTGCAC	CCTGGGGCCT	CCTCAGCCAA	TGGATGCCCT	200
G P S K	H C T	L G P	P Q P M	D A L	
D L Q	N T A P	W G L	L S Q	W M P W	
T F K	T L H	P G A S	S A N	G C P	
GGACTCTCCC	CTTCTTAGGA	CCTCTAGCAG	CTATAATATT	TTTACTCCTC	250
D S P	L L R T	S S S	Y N I	F T P L	
T L P	F L G	P L A A	I I F	L L L	
G L S P	S . D	L . Q	L . Y F	Y S S	
TTTGGACCCCT	GTATCTTCAA	CTTCCTTGTT	AAGTTTGTCT	CTTCCAGAAT	300
W T L	Y L Q	L P C .	V C L	F Q N	
F G P C	I F N	F L V	K F V S	S R I	
L D P	V S S T	S L L	S L S	L P E L	
TGAAGCTGTA	AAGCTACAAA	TAGTTCTTCA	AATGGAACCC	CAGATGCAGT	350
. S C K	A T N	S S S	N G T P	D A V	
E A V	K L Q I	V L Q	M E P	Q M Q S	
K L .	S Y K .	F F K	W N P	R C S	

10/27

FIG 5 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCATGACTAA	AATCTACCGT	GGACCCCTGG	ACCGGCCTGC	TAGACTATGC	400
H D .	N L P W	T P G	P A C	. T M L	
M T K	I Y R	G P L D	R P A	R L C	
P . L K	S T V	D P W	T G L L	D Y A	
TCTGATGTTA	ATGACATTGA	AGTCACCCCT	CCCGAGGAAA	TCTCAACTGC	450
. C .	. H .	S H P S	R G N	L N C	
S D V N	D I E	V T P	P E E I	S T A	
L M L	M T L K	S P L	P R K	S Q L H	
ACAACCCCTA	CTACACTCCA	ATTCACTAGG	AAGCAGTTAG	AGCAGTTGTC	500
T T P T	T L Q	F S R	K Q L E	Q L S	
Q P L	L H S N	S V G	S S .	S S C Q	
N P Y	Y T P	I Q .	E A V R	A V V	
AGCCAACCTC	CCCAACAGTA	CTTGGGTTTT	CCTGTTGAGA	GGGTGGACTG	550
A N L	P N S T	W V F	L L R	G W T E	
P T S	P T V	L G F S	C . E	G G L	
S Q P P	Q Q Y	L G F	P V E R	V D .	
AGAGACAGGA	CTAGCTGGAT	TTCCTAGGCT	GACTAAGAAT	CCCAAGCCT	600
R Q D	. L D	F L G .	L R I	P K P	
R D R T	S W I	S . A	D . E	S X S L	
E T G	L A G F	P R L	T K N	P X A X	
ANCTGGGAAG	GTGACCGCAT	CCATCTTTAA	ACATGGGGCT	TGCAACTTAG	650
X W E G	D R I	H L .	T W G L	Q L S	
X G K	V T A S	I F K	H G A	C N L A	
L G R	. P H	P S L N	M G L	A T .	
CTCACACCCG	ACCAATCAGA	GAGCTCACTA	AAATGCTAAT	CAGGCAAAAA	700
S H P	T N Q R	A H .	N A N	Q A K T	
H T R	P I R	E L T K	M L I	R Q K	
L T P D	Q S E	S S L	K C .	S G K N	

11/27

FIG 5 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAGGAGGTAA	AGCAATAGCC	AATCATCTAT	TGCCTGAGAG	CACAGCGGGA	750
G G K	A I A	N H L L	P E S	T A G	
Q E V K	Q . P	I I Y	C L R A	Q R E	
R R .	S N S Q	S S I	A . E	H S G K	
AGGACAAGGA	TTGGGATATA	AACTCAGGCA	TTCAAGCCAG	CAACAGCAAC	800
R T R I	G I .	T Q A	F K P A	T A T	
G Q G	L G Y K	L R H	S S Q	Q Q Q P	
D K D	W D I	N S G I	Q A S	N S N	
CCOCTTTGGG	TCCCTCCCA	TTGTATGGGA	GCTCTGTTTT	CACTCTATTT	850
P F G	S P P I	V W E	L C F	H S I S	
P L G	P L P	L Y G S	S V F	T L F	
P L W V	P S H	C M G	A L F S	L Y F	
CACTCTATTA	AATCATGCAA	CTGCACTCTT	CTGGTCCGTG	TTTTTTATGG	900
L Y .	I M Q	L H S S	G P C	F L W	
H S I K	S C N	C T L	L V R V	F Y G	
T L L	N H A T	A L F	W S V	F F M A	
CTCAAGCTGA	GCTTTTGTTC	GOCATCCACC	ACTGCTGTTT	GCCACCGTCA	950
L K L S	F C S	P S T	T A V C	H R H	
S S .	A F V R	H P P	L L F	A T V T	
Q A E	L L F	A I H H	C C L	P P S	
CAGACCCGCT	GCTGACTTCC	ATCCCTTTGG	ATCCAGCAGA	GTGTCCACTG	1000
R P A	A D F H	P F G	S S R	V S T V	
D P L	L T S	I P L D	P A E	C P L	
Q T R C	. L P	S L W	I Q Q S	V H C	
TGCTCCTGAT	CCAGCGAGGT	ACCCATTGCC	ACTCCCGATC	AGGCTAAAGG	1050
L L I	Q R G	T H C H	S R S	G . R	
C S .	S S E V	P I A	T P D Q	A K G	
A P D	P A R Y	P L P	L P I	R L K A	

12/27

FIG 5 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTTGCCATTG	TTCCTGCATG	GCTAAGTGCC	TGGGTTTGTC	CTAATAGAAC	1100
L A I V	P A W	L S A	W V C P	N R T	
L P L	F L H G	. V P	G F V	L I E L	
C H C	S C M	A K C L	G L S	. . N	
TGAACACTGG	TCACTGGGTT	CCATGGTTCT	CTTCCATGAC	CCACGGCTTC	1150
E H W	S L G S	M V L	F H D	P R L L	
N T G	H W V	P W F S	S M T	H G F	
. T L V	T G F	H G S	L P .	P T A S	
TAATAGAGCT	ATAACACTCA	CCGCATGGCC	CAAGATTCCA	TTCCTTGGTA	1200
I E L	. H S	P H G P	R F H	S L V	
. . S Y	N T H	R M A	Q D S I	P W Y	
N R A	I T L T	A W P	K I P	F L G I	
TCTGTGAGGC	CAAGAACCCC	AGGTCAGAGA	ANGTGAGGCT	TGCCACCATT	1250
S V R P	R T P	G Q R	X . G L	P P F	
L . G	Q E P Q	V R E	X E A	C H H L	
C E A	K N P	R S E X	V R L	A T I	
TGGGAAGTGG	CCCACTGCCA	TTTTGGTAGC	GGCCCACCAC	CATCTTGGGA	1300
G K W	P T A I	L V A	A H H	H L G S	
G S G	P L P	F W .	R P T T	I L G	
W E V A	H C H	F G S	G P P P	S W E	
GCTGTGGGAG	CAAGGATCCC	CCAGTAAACA			1329
C G S	K D P	P V T			
A V G A	R I P	Q .			
L W E	Q G S P	S N			

13/27

FIG 6

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCTAGAACGT	ATTCTGGAGA	ATTGGGACCA	ATGTGACACT	CAGACGCTAA	50
P R T Y	S G E L	G P M .	H S D A	K	
L E R I	L E N W	D Q C D	T Q T L	R	
. N V F	W R I G	T N V T	L R R .		
GAAAGAAACG	ATTTATATTC	TTCTGCAGTA	CCGCCTGGCC	ACAATATCCT	100
K E T I	Y I L L	Q Y R L	A T I S	S	
K K R F	I F F C	S T A W	P Q Y P		
E R N D	L Y S S	A V P P	G H N I	L	
CTTCAAGGGA	GAGAAACCTG	GCTTCTCTGAG	GGAAGTATAA	ATTATAACAT	150
S R E R	N L A S .	G K Y K	L . H		
L Q G R	E T W L	P E G S	I N Y N	I	
F K G E	K P G F	L R E V .	I I T S		
CATCTTACAG	CTAGACCTCT	TCTGTAGAAA	GGAGGGCAAA	TGGAGTGAAG	200
H L T A	R P L L .	K G G Q	M E . S		
I L Q L	D L F C	R K E G	K W S E	V	
S Y S .	T S S V	E R R A	N G V K		
TGCCATATGT	GCAAACTTTC	TTTTCATTAA	GAGACAACTC	ACAATTATGT	250
A I C A	N F L F	I K R Q	L T I M .		
P Y V Q	T F F S	L R D N	S Q L C		
C H M C	K L S F	H . E T	T H N Y	V	
AAAAAGTGIG	GTTTATGCCC	TACAGGAAGC	CCTCAGAGTC	CACCTCCCTA	300
K V W F	M P Y R	K P S E	S T S L		
K K C G	L C P T	G S P Q	S P P P	Y	
K S V V	Y A L Q	E A L R	V H L P	T	
CCCCAGGGTC	CCCTCCCCGA	CTCCTTCTCT	AACTAATAAG	GACCCCCCTT	350
P Q R P	L P D S	F L N . .	G P P F		
P S V .	P S P T	P S S T	N K D P	P L	
P A S P	P R L L	P Q L I	R T P L		
TAACCCAAAC	GGTCCAAAAG	GAGATAGACA	AAGGGGTAAA	CAATGAACCA	400
N P N G	P K G D	R Q R G	K Q . T	K	
T Q T V	Q K E I	D K G V	N N E P		
. P K R	S K R R .	T K G .	T M N Q		
AAGAGTGCCA	ATATTCCCCG	ATTATGCCCC	CTCCAAGCAG	TGAGAGGAGG	450
E C Q Y	S P I M	P P P S	S E R R		
K S A N	I P R L	C P L Q	A V R G	G	
R V P I	F P D Y	A P S K	Q . E E	E	
AGAATTCCGC	CCAGCCAGAG	TGCCTGTACC	TTTTTCTCTC	TCAGACTTAA	500
R I R P	S Q S A	C T F F	S L R L	K	
E F G P	A R V P	V P F S	L S D L	K	
N S A Q	P E C L	Y L F L	S Q T .		

14/27

FIG 6 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGCAAATTAA	AATAGACCTA	GGTAAATTCT	CAGATAACCC	TGACGGCTAT	550
A N .	N R P R .	I L R .	P .	R L Y	
Q I K	I D L	G K F S	D N P	D G Y	
S K L K	. T .	V N S	Q I T L	T A I	
ATTGATGTTT	TACAAGGGTT	AGGACAATCC	TTTGATCTGA	CATGGAGAGA	600
. C F	T R V	R T I L	. S D	M E R	
I D V L	Q G L	G Q S	F D L T	W R D	
L M F	Y K G .	D N P	L I .	H G E I	
TATAATGTTA	CTACTAAATC	AGACACTAAC	CCCAAATGAG	AGAAGTGCCG	650
Y N V T	T K S	D T N	P K .	E K C R	
I M L	L L N Q	T L T	P N E	R S A A	
. C Y	Y .	I R H .	P	Q M R E V P	
CTGTAAGTGC	AGCCCGAGAG	TTTGGCGATC	TTTGGTATCT	CAGTCAGGCC	700
C N C	S P R V	W R S	L V S	Q S G Q	
V T A	A R E	F G D L	W Y L	S Q A	
L .	L Q	P E S	L A I	F G I S V R P	
AACAATAGGA	TGACAACAGA	GGAAGAACA	ACTCCACAG	GCCAGCAGGC	750
Q .	D D N R	G K N N	S H R	P A G	
N N R M	T T E	E R T	T P T G	Q Q A	
T I G	. Q Q R	K E Q	L P Q	A S R Q	
AGTTCCCACT	GTAGACCCCTC	ATTGGGACAC	AGAATCAGAA	CATGGAGATT	800
S S Q C	R P S	L G H	R I R T	W R L	
V P S	V D P H	W D T	E S E	H G D W	
F P V	. T L	I G T Q	N Q N	M E I	
GGTGCCACAA	ACATTGCTA	ACTTGGGTGC	TAGAAGCACT	GAGGAAACT	850
V P Q	T F A N	L R A	R R T	E E N .	
C H K	H L L	T C V L	E G L	R K T	
G A T N	I C .	L A C	. K D .	G K L	
AGGAAGAAGC	CTATGAATTA	CTCAATGATG	TCCACTATAA	CACAGGGAAA	900
E E A	Y E L	L N D V	H Y N	T G K	
R K K P	M N Y	S M M	S T I T	Q G K	
G R S	L .	I T Q .	C P L .	H R E R	
GGAAGAAAT	CTTACTGCTT	TTCTGGACAG	ACTAAGGGAG	GCATTGAGGA	950
G R K S	Y C F	S G Q	T K G G	I E E	
E E N	L T A F	L D R	L R E	A L R K	
K K I	L L L	F W T D	. G R H .	G	
AGCATACCTC	CCTGTCACT	GACTCTATTG	AAGGCCAACT	AATCTTAAAG	1000
A Y L	P V T .	L Y .	R P T	N L K G	
H T S	L S P	D S I E	G Q L	I L K	
S I P P	C H L	T L L	K A N .	S . R	

FIG 6 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GATAAGTTTA	TCACTCAGTC	AGCTGCAGAC	ATTAGAAAAA	ACTTCAAAAG	1050
. V Y H S V	S C R H .	K K L Q K			
D K F I T Q S	A A D I R	K N F K S			
I S L S L S Q	L Q T L E	K T S K V			
TCTGCCTTAG	GCCCCGAGCA	GAACCTTAGAA	ACCCATATTTA	ACTTGGCATC	1100
S A L G P E Q	N L E T L	F N L A S			
L P . A R S R	T . K P Y	L T W H P			
C L R P G A	E L R N P	I . L G I			
CTCAGTTTTT	TATAATAGAG	ATCAGGAGGA	GCAGGCGAAA	CGGGACAAAC	1150
S V F Y N R D	Q E E Q A	K R D K R			
Q F F I I E	I R R S R	R N G T N			
L S F L . .	R S G G A	G E T G Q T			
GGGATAAAAA	AAAAAGGGGG	GGTCCACTAC	TTTAGTCATG	GCCCTCAGGC	1200
D K K K R G	G P L L .	S W P S G			
G I K K K G G	V H Y F S	H G P Q A			
G . K K K G G	S T T L V	M A L R Q			
AAGCAGACTT	TGGAGGCTCT	GCAAAAGGGA	AAAGCTGGGC	AAATCAAATG	1250
K Q T L E A L	Q K G K A	G Q I K C			
S R L W R L C	K R E K L	G K S N A			
A D F G G S	A K G K S	W A N Q M			
CCTAATAGGG	CTGGCTTCCA	GTGCGGTCTA	CAAGGACACT	TTAAAAAAGA	1300
L I G L A S S	A V Y K D	T L K K I			
. . G W L P	V R S T R	T L . K R			
P N R A G F Q	C G L Q G	H F K K D			
TTATCCAAGT	AGAAATAAGC	CGCCCCCTTG	TCATGCCCC	TTACGTCAAG	1350
I Q V E I S	R P L V H	A P Y V K			
L S K . K .	A A P L S	M P L T S R			
Y P S R N K P	P P C P C	P L R Q G			
GGAATCACTG	GAAGGCCCCAC	TGCCCCAGGG	GATGAAGATA	CTCTGAGTCA	1400
G I T G R P T	A P G D E	D T L S Q			
E S L E G P L	P Q G M K	I L . V R			
N H W K A H	C P R G .	R Y S E S			
GAGCCATTA	ACCAGATGAT	CCAGCAGCAG	GACTGAGGGT	GCCCCGGGCG	1450
K P L T R .	S S S R T	E G A R G E			
S H . P D D	P A A G L	R V P G A			
E A I N Q M I	Q Q Q D .	G C P G R			
AGGGCCAGCC	CATGCCATCA	CCCTCACAGA	GCCCCGGGTA	TGTTTGACCA	1500
R Q P M P S	P S Q S P	G Y V . P			
S A S P C H H	P H R A P	G M F D H			
A P A H A I T	L T E P R	V C L T I			

16/27

FIG 6 (suite)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TTGAGAGCCA	A			1511
L	R	A		
.	E	P		
E	S	Q		

17/27

FIG 7

10	20	40	50
1234567890 12 4	12 45 7 80 12 45 67890	1234567890	
ATGGGCAGCA GCA	TGTCATAC AGGAAAGGCC TGGTGCCGCG		50
M G S S	H H S S G L V P R		
CGGCAGCCAT ATG	TGACT GTGG ACAGCAATG GGTGGGATCC		100
G S H M A	T G G Q Q M G R I L		
TAGAAGTAT TCTG	TGGACCAAT GTGACACTCA GACGCTAAGA		150
E R I L E M W D Q C D T Q T L R			
AAGAAAGAT TTAT	CTCCAGTACC GCGTGCCAC AATATCTCT		200
K K R F I	C S T A W P Q Y P L		
TCAAGGGAGA GAA	TTTCTAACGG AAGTATAAAT TATAACATCA		250
Q G R E T	P E G S I N Y N I I		
TCCTACAGCT AGAC	TGTAGAAAGG AGGGCAAATG GAGTGAAGTG		300
L Q L D I	C R K E G K W S E V		
GCATATGTGC AAAC	TTTATTAGA GACAATCAC AATTATGTAA		350
P Y V Q T	S L E D N S Q L C K		
AAGTGTGGT TTAT	CAGGAGGCC TCGAGTCCA CCTCCCTACC		400
K C G L C	G S P Q S P P P Y P		
CCAGGGTCCC CTCC	CCTTCTCAA CTATAAGGA CCCCCCTTTA		450
S V P S E	P S S T N K D P P L		
AGCCAAAGG TCCAA	GATAGACAAA GGGTAAACA ATGAACCAAA		500
D O T V Q	E I D K G V N N E P K		
GAGTCCAT ATTC	TATGCCCCCT CCAAGCAGTG AGAGGAGGAG		550
S A N I P	C P L Q A V R G G E		
AATGGGCCC AGCC	CCTGTACCTT TTTCTCTCTC AGACTTAAAG		600
P G P A F	P V P F S L S D L K		
CAATTAATAA TAGA	TAAATTCTCA GATAACCTG ACGGCTATAT		650
Q T K I D	G K F S D N P D G Y I		
TGATGTTTTA CAAG	GACAATCCTT TGATCTGACA TGGAGAGATA		700
D V L Q G	Q S F D L T W R D I		
TATGTACT ACTA	ACACTAACC CAATGAGAG AAGTGCCGCT		750
M L L L	T L T P N E R S A A		
GTAACTGCAG CCG	TGCGTATT TGGTATCTCA GTCAGGCCAA		800
V T A A R	G D L W Y L S Q A N		
CAATAGGATG AC	AAAGACAC TCCACAGGC CAGCAGGCAG		850
N R M T T	E T T P T G Q Q A V		
TTCGAGTGT AGA	TGGACACAG AATCAGAACA TGGAGATTGG		900
P S V D E	W D T E S E H G D W		

18/27

FIG 7 (suite)

10	20	40	50	
1234567890	10 20 30 40 50	1234567890	1234567890	
TGCCACAAAC	ATG	TCGCTA	GAGGACTGA	GGAAAACTAG 950
C H K H		V	E G L R	K T R
GAAGAAGCCT	ATG	CAATGATTC	CACTATAACA	CAGGGAAAGG 1000
K K P M N		N M S T I T	Q G K E	
AAGAAAATCT	TACT	CTGACAGAC	TAAGGGAGGC	ATTGAGGAAG 1050
E N L T		L D R L	R E A	L R K
CATACCTGCC	TGTC	CTCTATTGAA	GGCCAACTAA	TCTTAAAGGA 1100
H T S L		S I E	G Q L I	L K D
TAAGTTTATC	ACTCA	CTCAGACAT	TAGAAAAAC	TTCAAAAGTC 1150
K F I T		A D I	R K N	F K S L
TGCTTAAGCT	TGCG	CTGAGCAC	ACCAACACA	CCACTGAGAT 1200
P K L A		L E H H	H H H	H D
CCGGCTGCTA	ACAA	AAAGGAAGCT	GAGTGGCIN	GTEGCA 1247
P A A N		K E A	E L A X	G

FIG 8

10	20	30	40	50	
1234567890 1234567890 1234567890 1234567890	0	10	20	30	
ATGGCTAGCA TGAC GATC ATGCTAGCA GGTGCGATCC TAGAACGTAT	50				
M A S M T G N Q L I G R I L E R I					
TCTGGAGAAT TGCG GATC GTGACGATA GACGCTAAGA AAGAAACGAT	100				
L E N W I L I T L R K K R F					
TTATATTCTT CTGG GATC GCTG GATC AATATCCTCT TCAAGGGAGA	150				
I F F C T T A W P Q Y P L Q G R					
GAAACCTGGC TTCTG GATC AACTATATAT TATAACATCA TCTTACAGCT	200				
E T W L P E G S I I Y N I I L Q L					
AGACCTCTTC TGTA GATC AGGC GATC GAGTGAAGTG CCATATGTGC	250				
D L F C E N E G R W S E V P Y V Q					
AAACTTTCTT TTCAATAGC GACAATGATC AATTATGTAA AAAGTGTGGT	300				
T F F S I E D N S Q L C K K C G					
TTATGCGCTA CAGG GATC TGGA GATC CCTCCCTACC CCAGCGTCCC	350				
G C P T G S I Q S I P P Y P S V P					
GTCCCCGACT CCTG GATC CTATATATA CCCCCCTTTA ACCCAAACGG	400				
S P T P S E T N I D P P L T Q T V					
TGCAAAAGGA GATACACAA GGGGTAAACA ATGAACCAAA GAGTGCCAAT	450				
Q K E I I K G V N N E P K S A N					
ATTCCCCGAT TATGCGCTT CCAAG GATC AGAGGAGGAG AATTGGGCCC	500				
L P R L C P L Q A V R G G E F G P					
AGCCAGAGTG CCTG GATC TTCTTTCTTC AGACTTAAAG CAAATTAAAA	550				
A R V P V P F S L S D L K Q I K I					
TGACCTAGG TAAATUTCA GATAACCTG ACGGCTATAT TGATGTTTTA	600				
L L G K F S D N P D G Y I D V L					
TGAGGGTTAG GACAATCTT TGATCTGACA TGGAGAGATA TAATGTTACT	650				
G G L G Q S F D L T W R D I M L L					
AGTAAATCAG AACTATGCC CAAATAGAG AAGTGCCGCT GTAACTGCAG	700				
N Q T L T P N E R S A A V T A A					
CTCCAGAGTT TGGG GATC TGGTATCTTA GTCAGGCCAA CAATAGGATG	750				
R E F G L L W Y L S Q A N N R M					
ACAACAGAGG AAGGATGATG TCCCATATC CAGCAGGCAG TTCCAGTGT	800				
T T E E R F T P T G Q Q A V P S V					
TGACCTCAT TGCG GATC AATCAGATA TGGAGATTGG TGCCACAAAC	850				
R H W D T E S E H G D W C H K H					
TAATGTTTAC TTGGTCTTA GAGGCTCTA GGAAACTAG GAAGAAGCCT	900				
T E T C I I E G L R K T R K K P					

10	0	40	50
1234567890 1234 5	12345 7 0	1234567890	1234567890
ATGAATTACT CCAI A	CACTA TTA A	CAGGGAAAGG AAGAAAATCT	950
M N Y S	T	Q G K E E N L	
TACTGCTTTT CTAI D	TAAGG AAC C	ATTGAGGAAG CATACTCC	1000
T A F L	I	L R K H T S L	
TGTCACCTGA CTCT TTA A	GGCA CTA A	TCTTAAAGGA TAAGTTTATC	1050
S P D S	G C L I	L K D K F I	
ACTCAGTCAG CTGG GAA C	TAGAA AAC	TTCAAAAGTC TGCTTAAGCT	1100
T Q S A A D	R K N	F K S L P K L	
TGGGGGCGCA CTGG GAA C	ACCAC AGCA	CCAGTACAT CCGGCTGCTA	1150
A A A L E H H	H H H	D P A A N	
ACAAAGCCCG AAG A	GACTTGGCTG	/	1186
K A R K	E L		

FIG 9

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTGCC	TCTCCCAATT	50
C P L C	S . S	S T G	A H C L	S Q L	
V R C	A P D P	A Q A	P I A	S P N W	
S A V	L L I	Q H R R	P L P	L P I	
GGGCTAAAGG	CTTGCCATTG	TTCTGCACA	GCTAAGTGCC	TGGGTTCATC	100
G . R	L A I V	P A Q	L S A	W V H P	
A K G	L P L	F L H S	. V P	G F I	
G L K A	C H C	S C T	A K C L	G S S	
CTAATCGAGC	TGAACACTAG	TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	150
N R A	E H .	S L G S	T V L	F H D	
L I E L	N T S	H W V	P R F S	S M T	
. S S	. T L V	T G F	H G S	L P . P	
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L	I E L	. H S	L H G P	R F H	
H G F	. . S Y	N T H	C M V	Q D S I	
M A S	N R A	I T L T	A W S	K I P	
TTCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E	S V R P	R T P	G Q R	T Q G L	
P W N	P . D	Q E P Q	V R E	H K A	
F L G I	R E T	K N P	R S E N	T R L	
TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTTGGAAGC	AGCCCGCCAC	300
P P C	W K Q	P T T I	L E A	A R H	
C H H V	G S S	P P P	F W K Q	P A T	
A T M	L E A A	H H H	F G S	S P P L	
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	AGGTAACAAT	TTGGTGACCA	350
Y L G S	S G S	K D P	R . Q	F G D H	
I L G	A L G A	R T P	G N N	L V T T	
S W E	L W E	Q G P Q	V T I	W . P	
CGAAGGGACC	TGAATCCGCA	ACCATGAAGG	GATCTCCAAA	GCAATTGGAA	400
E G T	. I R N	H E G	I S K	A I G N	
K G P	E S A	T M K G	S P K	Q L E	
R R D L	N P Q	P . R	D L Q	S N W K	

FIG 9 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGTTCTCTCC	CAAGGCAAAA	ATGCCCCCTAA	GATGTATTCT	GGAGAATTGG	450
V P P	K A K	M P L R	C I L	E N W	
M F L P	R Q K	C P .	D V F W	R I G	
C S S	Q G K N	A P K	M Y S	G E L G	
GACCAATTTG	ACCTCAGAC	AGTAAGAAAA	AAATGACTTA	TATTCTTCTG	500
D Q F D	P Q T	V R K K	. L I	F F C	
T N L	T L R Q	. E K	N D L	Y S S A	
P I .	P S D	S K K K	M T Y	I L L	
CAGTACCGCC	CTGGCCAAGA	TATCTCTCTC	AAGGGGGAGA	AACCTGGCCT	550
S T A	L A T I	S S S	R G R	N L A S	
V P P	W P R	Y P L Q	G G E	T W P	
Q Y R P	G H D	I L F	K G E K	P G L	
CCTGAGGGAA	GTATAAATTA	TAACACCATC	TTACAGCTAG	ACCTGTTTTG	600
. G K	Y K L	. H H L	T A R	P V L	
P E G S	I N Y	N T I	L Q L D	L F C	
L R E V	. I I	T P S	Y S .	T C F V	
TAGAAAAGGA	GGCAAATGGA	GTGAAGTGCC	ATATTACAA	ACTTTCTTTT	650
. K R R	Q M E	. S A	I F T N	F L F	
R K G	G K W S	E V P	Y L Q	T F F S	
E K E	A N G	V K C H	I Y K	L S F	
CATTAAAAGA	CAACTCGCAA	TTATGTTAAC	AGTGTGATTT	GTGTTCTTAC	700
I K R	Q L A I	M L T V	. F	V F L H	
L K D	N S Q	L C .	Q	C D L C S Y	
H . K T	T R N	Y V N	S V I C	V P T	
ACGGAAGCCC	TCAGATTCTA	CTCCCCACCC	CCGGCATCTC	CCCTGAATCC	750
G S P	Q I L	L P T P	G I S	P E S	
T E A L	R F Y	S P P	P A S P	L N P	
R K P	S D S T	P H P	R H L	P . I P	
CTCCCCAACT	TATT				764
L P N L					
S P T Y					
P Q L I					

FIG 10

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTTGCC	TCTCCCAATT	50
C P L C	S . S	S T G	A H C L	S Q L	
V R C	A P D P	A Q A	P I A	S P N W	
S A V	L L I	Q H R R	P L P	L P I	
GGGCTAAAGG	CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTTCATC	100
G . R	L A I V	P A Q	L S A	W V H P	
A K G	L P L	F L H S	. V P	G F I	
G L K A	C H C	S C T	A K C L	G S S	
CTAATCGAGC	TGAACACTAG	TCACTGGGGTT	CCACGGTTCT	CTTCCATGAC	150
N R A	E H .	S L G S	T V L	F H D	
L I E L	N T S	H W V	P R F S	S M T	
. S S	. T L V	T G F	H G S	L P . P	
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L	I E L	. H S	L H G P	R F H	
H G F	. . S Y	N T H	C M V	Q D S I	
M A S	N R A	I T L T	A W S	K I P	
TTCCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E	S V R P	R T P	G Q R	T Q G L	
P W N	P . D	Q E P Q	V R E	H K A	
F L G I	R E T	K N P	R S E N	T R L	
TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTTGGAAGC	GGCCCGCCAC	300
P P C	W K Q	P T T I	L E A	A R H	
C H H V	G S S	P P P	F W K R	P A T	
A T M	L E A A	H H H	F G S	G P P L	
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	CAGGTAACAA	TTTGGTGACC	350
Y L G S	S G S	K D P	Q V T I	W . P	
I L G	A L G A	R T P	R . Q	F G D H	
S W E	L W E	Q G P P	G N N	L V T	
ACGAAGGGAC	CTGAATCCGC	AACCATGAAG	GGATCTCCAA	AGCAATTGGA	400
R R D	L N P Q	P . R	D L Q	S N W K	
E G T	. I R	N H E G	I S K	A I G	
T K G P	E S A	T M K	G S P K	Q L E	

FIG 10 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AATGTTCCCTC	CCAAGGCAAA	AATGCCCCCTA	AGATGTATTC	TGGAGAATTG	450
C S S	Q G K	N A P K	M Y S	G E L	
N V P P	K A K	M P L	R C I L	E N W	
M F L	P R Q K	C P .	D V F	W R I G	
GGACCAATCT	GACCCTCAGA	CAGTAAGAAA	AAAAATGACT	TATATTCTTC	500
G P I .	P S D	S K K	K N D L	Y S S	
D Q S	D P Q T	V R K	K M T	Y I L L	
T N L	T L R	Q .	E K	K . L I F F	
TGCAGTACCG	CCTGGCCACG	GATATCCTCT	TCAAGGGGGA	GAAACCTGGC	550
A V P	P G H G	Y P L	Q G G	E T W P	
Q Y R	L A T	D I L F	K G E	K P G	
C S T A	W P R	I S S	S R G R	N L A	
CTCCTGAGGG	AAGTATAAAT	TATAACACCA	TCTTACAGCT	AGACCTGTTT	600
P E G	S I N	Y N T I	L Q L	D L F	
L L R E	V .	I I T P	S Y S .	T C F	
S .	G K Y K L	. H H	L T A	R P V L	
TGTAGAAAAG	GAGGCAAATG	GAGTGAAGTG	CCATATTTAC	AAACTTTCTT	650
C R K G	G K W	S E V	P Y L Q	T F F	
V E K	E A N G	V K C	H I Y	K L S F	
. K R	R Q M E .	S A	I F T	N F L	
TTCATTAAAA	GACAACTCGC	AATTATGTAA	ACAGTGTGAT	TTGTGTCCCTA	700
S L K	D N S Q	L C K	Q C D	L C P T	
H .	K T T R	N Y V N	S V I	C V L	
F I K R	Q L A	I M .	T V .	F V S Y	
CAGGAAGCCC	TCAGATCTAC	CTCCCTACCC	CGGCATCTCC	CTGACTCCTT	750
G S P	Q I Y	L P T P	A S P .	L L	
Q E A L	R S T	S L P	R H L P	D S F	
R K P	S D L P	P Y P	G I S	L T P S	
CCCCAACTAA	TAAGGACCCA	CTTCAGCCCA	AACAGTCCAA	AAGGACATAG	800
P Q L I	R T H	F S P	N S P K	G H	
P N .	. G P T	S A Q	T V Q	K D I	
P T N	K D P	L Q P K	Q S K	R T .	

FIG 11

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GGCATTGATA	GCACCCATCA	GATGGCCAAA	TCATTATTTA	CTGGACCAGG	50
G I D S	T H Q	M A K	S L F T	G P G	
A L I	A P I R	W P N	H Y L	L D Q A	
H . .	H P S	D G Q I	I I Y	W T R	
CCTTTTCAAA	ACTATCAAGC	AGATAGGGCC	CGTGAAGCAT	GCCAAAGAAA	100
L F K	T I K Q	I G P	V K H	A K E I	
F S K	L S S R	. G P	. S M	P K K	
P F Q N	Y Q A	D R A	R E A C	Q R N	
TAATCCCCTG	CCTTATCGCC	ATGTTCTTTC	AGGAGAACAA	AGAACAGGCC	150
I P C	L I A	M F L Q	E N K	E Q A	
. S P A	L S P	C S F	R R T K	N R P	
N P L	P Y R H	V P S	G E Q	R T G H	
ATTACCCAGG	GGAAGACTGG	CAACTAGATT	TTACCCACAT	GGCCAAATGT	200
I T Q G	K T G	N . I	L P T W	P N V	
L P R	G R L A	T R F	Y P H	G Q M S	
Y P G	E D W	Q L D F	T H M	A K C	
CAGGGATTTC	AGCATCTACT	AGTCTGGGCA	GATACTTTCA	CTGGTTGGGT	250
R D F	S I Y .	S G Q	I L S	L V G W	
G I S	A S T	S L G R	Y F H	W L G	
Q G F Q	H L L	V W A	D T F T	G W V	
GGAGTCTTCT	CCTTGTAGGA	CAGAAAAGAC	CCAAGAGGTA	ATAAAGGCAC	300
S L L	L V G	Q K R P	K R .	. R H	
G V F S	L . D	R K D	P R G N	K G T	
E S S	P C R T	E K T	Q E V	I K A L	
TAATGAAATA	ATTCCAGAT	TTGGACTTCC	CCCAGGATTA	CAGGGTGACA	350
. . N N	S Q I	W T S	P R I T	G . Q	
N E I	I P R F	G L P	P G L	Q G D N	
M K .	F P D	L D F P	Q D Y	R V T	

FIG 11 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCCGC	TTTCAAGGCT	GCAGTAACCC	AGGGAGTATC	CCAGGTGTTA	400
W P R	F Q G C	S N P	G S I	P G V R	
G P A	F K A	A V T Q	G V S	Q V L	
M A P L	S R L Q	. P	R E Y P	R C .	
GGCATACAAT	ATCACTTACA	CTGTGCCTGG	AGGCCACAAT	CCTCCAGAAA	450
H T I	S L T	L C L E	A T I	L Q K	
G I Q Y	H L H	C A W	R P Q S	S R K	
A Y N	I T Y T	V P G	G H N	P P E K	
AGTCAAGAAA	ATGAATGAAA	CACTCAAAGA	TCTAAAAAAG	CTAACCCAAG	500
S Q E N	E . N	T Q R	S K K A	N P R	
V K K	M N E T	L K D	L K K	L T Q E	
S R K	. M K	H S K I	. K S	. P K	
AAACCCACAT	TGCATGACCT	GTTCTGTTGC	CTATAACCTT	ACTAAGAATC	550
N P H	C M T C	S V A	Y N L	T K N P	
T H I	A . P	V L L P	I T L	L R I	
K P T L	H D L	F C C	L . P Y	. E S	
CATAACTATC	CCCCAAAAAG	CAGGACTTAG	CCCATACGAG	ATGCTATATG	600
. L S	P K K	Q D L A	H T R	C Y M	
H N Y P	P K S	R T .	P I R D	A I W	
I T I	P Q K A	G L S	P Y E	M L Y G	
GATGGCCTTT	CCTAACCAAT	GACCTTGIGC	TTGACTGAGA	AATGGCCAAC	650
D G L S	. P M	T L C	L T E K	W P T	
M A F	P N Q .	P C A	. L R	N G Q L	
W P F	L T N	D L V L	D . E	M A N	
TTAGTTGCAG	ACATCACCTC	CTTAGCCAAA	TATCAACAAG	TTCTTAAAC	700
. L Q	T S P P	. P N	I N K	F L K H	
S C R	H H L	L S Q I	S T S	S . N	
L V A D	I T S	L A K	Y Q Q V	L K T	

27/27

FIG 11 (suite)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATCACAGGGA	ACCTGTCCCC	GAGAGGAGGG	AAAGGAACTA	TTCCACCCCTG	750
H R E P V P	E R R E	R N Y S T L			
I T G N L S P	R G G K G T I	P P W			
S Q G T C P R	E E G K E L	F H P G			
GTGACATG					758
V T					
. H					
D M					